Least-False and Local Misspecification Methods for Longitudinal Data with Dropout

Amal A. Almohisen

Thesis submitted for the degree of
Doctor of Philosophy

School of Mathematics & Statistics
Newcastle University
Newcastle upon Tyne
United Kingdom

March 2017

To my parents,
The reason of what I become today.
Thanks for your great support and continues care.

Acknowledgements

I would like to express deepest gratitude to my supervisor Prof. Robin Henderson for his full support, expert guidance, understanding and encouragement throughout my study and research. Without his incredible patience and timely wisdom and counsel, my thesis work would have been a frustrating and overwhelming pursuit.

I would also like to thank Dr. Jian, Prof. Shuhrat, and Dr. Arwa for helping me with my academic research during my PhD. study.

Thanks also go to King Saud University, Statistics and OR Department and External Joint Supervision Program staff for their support in facilitating my needs to complete my research. Also, special thanks go to my numerous friends who helped me throughout this academic exploration.

Finally, I would like to thank my husband, parents and sisters for their unconditional support during the last five years; I would not have been able to complete this thesis without their continuous love and encouragement.

Abstract

In any longitudinal study, a dropout before the final timepoint can rarely be avoided. The chosen dropout model is commonly one of these types: Missing Completely at Random (MCAR), Missing at Random (MAR), Missing Not at Random (MNAR) and Shared Parameter (SP). In this thesis, we present methods to estimate the longitudinal model parameters under a variety of different dropout models. These methods are Complete Case analysis (CC), Observed data analysis (Obs), Inverse Probability Weighted estimating equations (IPW), Linear Mixed Effect models (LME), Linear Increment models (LI) and Last Observation Carried Forward (LOCF). We estimate the parameters of the longitudinal model under MCAR, MAR, MNAR and SP for both simulated data and real data assuming two and three timepoint examples. We show that all methods work under the MCAR model as expected. Also, the LI method give consistent estimate under the SP model. The IPW and LME give consistent estimate under MAR, while no method work under MNAR.

We investigate the consequences of misspecifying the missingness mechanism by deriving the so called least false values. These are the values the parameter estimates converge to, when the assumptions may be wrong. This constitutes the central part of the thesis. In order to calculate the least false values, we use the approximation to the extended skew normal distribution (ESN) as produced in Ho et al. [2012]. We give closed form expressions to calculate the least false values β_3^* and β_4^* for LI, CC and LME methods. For the IPW, we provide a closed form for β_3^* under SP, MAR and MNAR while for β_4^* we failed to find closed form under MNAR and we use a numerical calculation instead. The knowledge of the least false values allows us to conduct sensitivity analysis which will be illustrated.

This method provides an alternative to a local misspecification sensitivity procedure which has been developed for likelihood-based analysis. The LME method is a likelihood based method, and this idea can be also adapted for the IPW estimating equation approach. We compare the results obtained by our method with the results found by using the local misspecification method. We show that Copas and Eguchi [2005] method and LME least false match very well. Both gave very close results. This suggests that our least false method can provide a credible alternative to Copas and Eguchi in sensitivity analysis. In fact it might be preferred since there is no assumption of local misspecification. Moreover, we apply the local misspecification and least false methods to estimate the bias and sensitivity for two real data examples with two timepoint and three timepoint data. We show how the IPW method is much more sensitive to misspecification than the LME method.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The problem of analyzing incomplete longitudinal data has been tackled by many authors. Some of the early references include Wu and Carroll [1988], Heyting et al. [1992], Diggle and Kenward [1994] and Schluchter [1992]. Earlier work on missing measurements was largely focused on forming algorithms for computational solutions (Afifi and Elashoff [1966]), but data imputation with the aid of powerful computers has compacted the problem to some extent. For later methodologies like the expectation-maximization (EM) see Dempster et al. [1977]. In this thesis, we discuss the concept of missing data mechanisms and present some methods to handle longitudinal data with dropout. The dropout model is commonly one of the following types: Missing Completely at Random (MCAR), Missing at Random (MAR), Missing Not at Random (MNAR) and Shared Parameter (SP) as introduced by Little and Rubin [2002] for the first three types and Follmann and Wu [1995] for shared parameter model. We will consider six general strategies for handling missing data : 1) Complete Case analysis (CC), 2) Observed data analysis (Obs), 3) Inverse Probability Weighted Estimating Equations with Missing at Random Assumption (IPW), 4) Linear Mixed Effect models (LME), 5) Linear Increment models (LI) and 6) Last Observation Carried Forward (LOCF). In Chapter 2, we estimate the parameters of the longitudinal model under MCAR, MAR, MNAR and SP for both simulated data and real data assuming, for simplicity, that there are only either two or three scheduled measurements times. We will refer to these as two timepoint and three timepoint situations respectively.

Next, we investigate the consequences of misspecifying the missingness mechanism by deriving the so called least false values (see Claeskens and Hjort [2008]). These are the values the parameter estimates converge to, when the assumptions may be wrong. This constitutes the central part of the thesis. We describe this method in detail for the Linear Increment method (LI) in Chapter 3, to see the performance under shared parameter, MAR and MNAR dropout models. Then, in Chapter 4, we apply the same procedure to investigate the IPW method. In Chapters 5 and 6 we investigate the CC and LME methods respectively. We adopt the procedures used by Diggle et al. [2007] and Ho et al.

[2012] to our setting in order to calculate the least false values.

The knowledge of the least false values allows us to conduct sensitivity analysis which will be illustrated in Chapters 7 and 8. Copas and Eguchi [2005] give a formula to estimate the bias under such misspecification using a likelihood approach. Their method is called a local misspecification sensitivity procedure. The LME is a likelihood based method, and this idea can be also adapted for the IPW estimating equation approach. We compare the results obtained by our method with the results found by using the local misspecification method. Moreover, we apply the local misspecification and least false methods to estimate the bias for two real data examples.

1.1   Assumptions

Longitudinal data is defined as information on a set of individuals that is collected at multiple follow up times. The statistical analysis aims to model the evolution of the response variable Y and investigate the relationship with the covariates.

Suppose there are n individuals in a study and each provides longitudinal responses Y and dropout information R. Generally we will assume a linear model for Y (in the absence of dropout) and logistic models for the probability of *continuing* to the next timepoint $t+1$ given that a subject is still under observation at time t. At times we will refer to a *true* or *generating* model as the way in which data are obtained, and to an *assumed* or *fitting* model as that chosen by the analyst for estimation.

For simplicity in this work, we will assume there are either two or three observations or treatment periods. The methods are of course more general.

For the most parts we will consider the simple case for either two or three timepoints. Assume to begin with that there are just two scheduled measurements. At time 1, there is a measurement provided for all subjects, denoted Y_{i1} for subject i. Then at time 2 some subjects drop out before measurement. Let $R_i=1$ indicate that we have a measurement at time 2 and $R_i=0$ otherwise. Let $Y_i=(Y_{i1},Y_{i2})^T$ and assume $E[Y_i]=x_i\beta$ where β is a parameter vector of dimension p and x_i is the design matrix associated with subject i, which is of dimension $2 \times p$. Our standard model assumes just one covariate and is:

$$\left.\begin{array}{l} Y_{1i} = \beta_1^G + \beta_2^G x_i + U_i + \epsilon_{1i} \\[2ex] Y_{2i} = \beta_3^G + \beta_4^G x_i + U_i + \epsilon_{2i} \end{array}\right\} Y_i = x_i\beta^G + U_i\mathbf{1} + \epsilon_i, \qquad (1.1)$$

where $Y_i = \begin{pmatrix} Y_{1i} \\ Y_{2i} \end{pmatrix}$, $x_i = \begin{pmatrix} 1 & x_i & 0 & 0 \\ 0 & 0 & 1 & x_i \end{pmatrix}$, $\beta^G = (\beta_1^G, \beta_2^G, \beta_3^G, \beta_4^G)^T$, $\mathbf{1} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$,

$$\epsilon_i = \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix}, \text{ and } x_i \sim N(0, \sigma_x^2) \quad U_i \sim N(0, \sigma_U^2), \quad \epsilon_{1i} \sim N(0, \sigma_{\epsilon_{1i}}^2), \epsilon_{2i} \sim N(0, \sigma_{\epsilon_{2i}}^2).$$

Also we have $i = 1, \ldots, n$ and the superscript G is used to indicate the true generating values. All subjects are observed at time 1, but some may dropout by time 2.

At times, we will write

$$\left. \begin{aligned} Y_{1i} &= \beta_1^G + \beta_2^G x_i + Z_{1i} \\ Y_{2i} &= \beta_3^G + \beta_4^G x_i + Z_{2i} \end{aligned} \right\} \quad Y_i = x_i \beta^G + Z_i, \tag{1.2}$$

where $Z_{i1} \sim N(0, \sigma_1^2)$, $Z_{i2} \sim N(0, \sigma_2^2)$ and $corr\,(Z_{i1}, Z_{i2}) = \rho$. Clearly $\rho = \sigma_U^2/\sigma_1\sigma_2$, $\sigma_1 = \sqrt{\sigma_U^2 + \sigma_{\epsilon_1}^2}$ and $\sigma_2 = \sqrt{\sigma_U^2 + \sigma_{\epsilon_2}^2}$.

We will also consider three timepoints. At time 1, there is a measurement provided for all subjects, denoted Y_{i1} for subject i, and then at time 2 some subjects drop out before this measurement. Let $R_{1i}=1$ indicate that we have a measurement at time 2 and $R_{1i}=0$ otherwise. Then, if a subject provides a measurement at time 2 let $R_{2i}=1$ indicate that we have a measurement at time 3 and $R_{2i}=0$ otherwise. Let $Y_i=(Y_{i1}, Y_{i2}, Y_{i3})^T$ and assume $E[Y_i]=x_i\beta$ where β is the p dimension parameter vector and x_i is now a $3 \times p$ design matrix associated with subject i. Usually, as above, we have a single covariate and so $p = 2$, including the intercept term.

1.2 Mechanisms of Missingness

Returning to the general case, the influence of missing data depends on the missingness mechanism, that is the probability model for missingness. Knowing the reason for the missingness is obviously helpful to handle missing data. We consider four general *missingness mechanisms* as introduced in Little and Rubin [2002] and Wu and Carroll [1988]. Assume R_i to be a response indicator vector for the ith subject so that $R_{ij}=1$ if Y_{ij} is observed, $R_{ij}=0$ if Y_{ij} is missing, where i and j refer to the jth observation for the ith subject.

We partition the complete data vector $Y_i = (Y_{iO}, Y_{iM})$ into those observed components, Y_{iO}, and those that are not observed Y_{iM}. Note that in the case of *dropout* then once a value is missing then so too are all later ones (monotone dropout). In this case $R = (1, 1, \cdots, 1, 0, 0, \ldots, 0)^T$. The alternative is *intermittent missingness* which allows observation to resume after missingness occurs. In this thesis we will concentrate on monotone dropout.

As in Bell and Fairclough [2013], we suppose there are n independent patients with m

scheduled measurements, and $m_i \leq m$ observed values for patient i which means we initially assume that all patients are observed. Let x_i represent covariates (such as treatment assignment). Without loss of generality, we can exclude x_i in the notation at times. Then the likelihood of the complete data is $f(Y_O, Y_M, R) = f(Y_O, Y_M) f(R|Y_O, Y_M)$ and of the observed data is

$$f(Y_O, R) = \int f(Y_O, Y_M) f(R|Y_O, Y_M) dY_M. \tag{1.3}$$

1.2.1   Missing completely at random dropout model (MCAR)

A variable is missing completely at random if the probability of missingness does not depend on outcomes Y, either Y_O or Y_M. For example, if each participant decides if he or she will answer for instance an education level question by flipping a coin and refusing to answer if a head shows up. If data are missing completely at random, then removing those cases with missing data does not influence or bias inferences. Formally, data are MCAR if

$$P(R_{ij} = 1|Y_{iO}, Y_{iM}, x_i) = P(R_{ij}|x_i)$$

so that the missingness is unrelated to the response data, whether its values are observed or are not observed. The missingness can depend on x_i. In terms of (1.3), under MCAR we have $f(R|Y_O, Y_M) = f(R)$ and so $f(Y_O, R) = f(R) \int f(Y_O, Y_M) dY_M = f(R) f(Y_O)$. If there are no shared parameters in $f(R)$ and $f(Y_O)$ we can ignore $f(R)$ and estimate from $f(Y_O)$ for the observed data only, without loss, except perhaps in efficiency.

The MCAR model that we will assume for the two timepoints example will be

$$\pi_i(\theta) = P(R_i = 1|Y_{1i}, Y_{2i}) = \frac{e^{\theta_0}}{1 + e^{\theta_0}} = \text{expit}(\theta_0) \tag{1.4}$$

where where, in general $\text{expit}(z) = e^z/(1 + e^z)$. As used, R_i is an indicator of observation ($R_i = 1$) or dropout ($R_i = 0$) at time 2. We have a similar model for three timepoints, with two potential missing values.

1.2.2   Missing at random dropout model (MAR)

Missing at random occurs when the probability of dropout depends on what available information there is. Thus, for any survey, if all the subjects answer such questions as sex, age, earnings, and ethnicity, then a response is considered missing at random if on answering a question such as education, the probability of not answering this question depends only on the other fully recorded variables.

In our case, data are MAR if $P(R_{ij} = 1|Y_{iO}, Y_{iM}, x_i) = P(R_{ij}|Y_{Oi}, x_i)$, hence the dropout

may depend on observed values. Turning to (1.3), under MAR we have

$$
\begin{aligned}
f(Y_O, R) &= \int f(Y_O, Y_M) f(R|Y_O) dY_M \\
&= f(R|Y_O) \int f(Y_O, Y_M) dY_M \\
&= f(R|Y_O) f(Y_O).
\end{aligned}
$$

Hence the joint density of observables can be modelled correctly. Again, if there are no shared parameters then we can ignore $f(R|Y_O)$ and estimate consistently from $f(Y_O)$ using maximum likelihood. Seaman et al. [2013] distinguish between realised MAR and everywhere MAR. The first means $P(R_{ij} = r|Y_{iO}, Y_{iM}) = P(R_{ij} = r|Y_{Oi})$ for the particular r seen in the data. The second means $P(R_{ij}|Y_{iO}, Y_{iM}) = P(R_{ij} = r|Y_{Oi})$ for every possible R. The second is a much stronger assumption. Mealli and Rubin [2015] make a similar point.

For simplicity in our investigations, we assume that the parameters are common between timepoints. Furthermore, we assume that Y_1 does not affect the probability of dropout after time 2. Both assumptions can be relaxed if necessary. Let the dropout parameters $\theta^M = (\theta_0^M, \theta_1^M)$. At the two timepoints situation, the MAR dropout logistic model is:

$$
\pi_i(\theta^M) = P(R_i = 1|Y_{1i}, Y_{2i}) = \frac{e^{\theta_0^M + \theta_1^M Y_{1i}}}{1 + e^{\theta_0^M + \theta_1^M Y_{1i}}} = \text{expit}(\theta_0^M + \theta_1^M Y_{1i}) \tag{1.5}
$$

And the three timepoints MAR version is

$$
\left.
\begin{aligned}
p_{1i}(\theta^M) = P(R_{1i} = 1|Y_{1i}, Y_{2i}, Y_{3i}) &= \frac{e^{\theta_0^M + \theta_1^M Y_{1i}}}{1 + e^{\theta_0^M + \theta_1^M Y_{1i}}} &= \text{expit}(\theta_0^M + \theta_1^M Y_{1i}) \\
\\
p_{2i}(\theta^M) = P(R_{2i} = 1|Y_{1i}, Y_{2i}, Y_{3i}, R_{1i} = 1) &= \frac{e^{\theta_0^M + \theta_1^M Y_{2i}}}{1 + e^{\theta_0^M + \theta_1^M Y_{2i}}} &= \text{expit}(\theta_0^M + \theta_1^M Y_{2i})
\end{aligned}
\right\} \tag{1.6}
$$

Note that we have used $p(.)$ for continuation probability, because in general this is conditional on not previously dropping out. The unconditional observation probabilities are

$$
\pi_{1i}(\theta) = p_{1i}(\theta) \quad \text{and} \quad \pi_{2i}(\theta) = p_{1i}(\theta) \times p_{2i}(\theta).
$$

1.2.3 Missing not at random dropout model (MNAR)

The missingness is called missing not at random, if it depends on unrecorded information which predicts the missing values. An example is that a patient was unsatisfied with a particular treatment thus this patient is more likely to quit the study. This missingness is not at random (unless that feeling of to be unsatisfied, is measured and observed for all patients). If missingness is not at random, we expect some bias in inferences.

Notationally, the missingness depends on Y_M, i.e. $P(R_{ij} = 1|Y_{iO}, Y_{iM}, x_i) \neq P(R_{ij} = 1|Y_{Oi}, x_i)$. Under MNAR, there is no simplification of (1.3) and so sometimes MNAR is also known as non-ignorable. Models can be true, but the observed data alone is not enough to check the model.

Let the dropout parameters be $\theta^{MN} = (\theta_0^{MN}, \theta_1^{MN})$. The MNAR version for the two timepoints example is the logistic model:

$$
\begin{aligned}
\pi_i(\theta^{MN}, \theta_2^{MN}) &= P(R_i = 1|Y_{1i}, Y_{2i}) = \frac{e^{\theta_0^{MN}+\theta_1^{MN}Y_{1i}+\theta_2^{MN}Y_{2i}}}{1 + e^{\theta_0^{MN}+\theta_1^{MN}Y_{1i}+\theta_2^{MN}Y_{2i}}} \\
&= \text{expit}(\theta_0^{MN} + \theta_1^{MN}Y_{1i} + \theta_2^{MN}Y_{2i}).
\end{aligned}
\tag{1.7}
$$

And for the three timepoints example, our MNAR logistic model is:

$$
\left.
\begin{aligned}
p_{1i}(\theta^{MN}, \theta_2^{MN}) = P(R_{1i} = 1|Y_{1i}, Y_{2i}.Y_{3i}) \quad &= \quad \frac{e^{\theta_0^{MN}+\theta_1^{MN}Y_{1i}+\theta_2^{MN}Y_{2i}}}{1+e^{\theta_0^{MN}+\theta_1^{MN}Y_{1i}+\theta_2^{MN}Y_{2i}}} \\
\\
&= \quad \text{expit}(\theta_0^{MN} + \theta_1^{MN}Y_{1i} + \theta_2^{MN}Y_{2i}) \\
\\
p_{2i}(\theta^{MN}, \theta_2^{MN}) = P(R_{2i} = 1|Y_{1i}, Y_{2i}.Y_{3i}, R_{1i} = 1) \quad &= \quad \frac{e^{\theta_0^{MN}+\theta_1^{MN}Y_{2i}+\theta_2^{MN}Y_{3i}}}{1+e^{\theta_0^{MN}+\theta_1^{MN}Y_{2i}+\theta_2^{MN}Y_{3i}}} \\
\\
&= \quad \text{expit}(\theta_0^{MN} + \theta_1^{MN}Y_{2i} + \theta_2^{MN}Y_{3i})
\end{aligned}
\right\}
\tag{1.8}
$$

Again,
$$
\pi_{1i}(\theta) = p_{1i}(\theta, \theta_2) \quad \text{and} \quad \pi_{2i}(\theta, \theta_2) = p_{1i}(\theta, \theta_2) \times p_{2i}(\theta, \theta_2).
$$

1.2.4 Shared parameter dropout model (SP)

In shared parameter models a random effect U is shared between the repeated measures model and the missing data mechanism model, where U is an unobserved subject specific random effect. The shared parameter dropout logistic model for the two timepoints example is defined as:

$$
P(R_i = 1|Y_{1i}, Y_{2i}, U_i) = \frac{e^{\theta_0^{SP}+\theta_1^{SP}U_i}}{1 + e^{\theta_0^{SP}+\theta_1^{SP}U_i}} = \text{expit}(\theta_0^{SP} + \theta_1^{SP}U_i).
\tag{1.9}
$$

Typically U_i will affect the distributions of Y_{1i} and Y_{2i} also.

Similarly, the shared parameter dropout logistic model for three timepoints example will be:

$$
P(R_i = 1|Y_{1i}, Y_{2i}, Y_{3i}, U_i) = \frac{e^{\theta_0^{SP}+\theta_1^{SP}U_i}}{1 + e^{\theta_0^{SP}+\theta_1^{SP}U_i}} = \text{expit}(\theta_0^{SP} + \theta_1^{SP}U_i).
\tag{1.10}
$$

We use θ^{SP}, θ^{M} and θ^{MN} to denote values under SP, MAR and MNAR respectively.

1.3   Thesis Outline

The structure of the thesis is as follows. Having introduced the missing data mechanisms, we consider four dropout models: MCAR, MAR, MNAR and SP in subsequence chapters. In Chapter 2, we will explore six dropout modelling methods: CC, Obs, IPW, LME, LI and LOCF. We will estimate the regression parameters for both simulated and real data using each of the methods under different dropout models, then in the following chapters we will study some of them in more detail. In Chapter 3, we will see how the LI method performs under SP, MAR and MNAR dropout by calculating the least false values. The performance and the least false values of IPW, CC and LME methods under different dropout modelling will be discussed and calculated in Chapters 4, 5 and 6 respectively. In Chapter 7, we will conduct a sensitivity analysis to see the effect of local misspecification of the dropout model when using the LME method under the MAR assumption. We will compare our results found by using the least false method with the local misspecification method as introduced by Copas and Eguchi [2005]. We will apply both methods on simulated and real data examples with two and three timepoints. A similar sensitivity analysis will be used in Chapter 8 for the IPW method. Chapter 9 will sum up the most important ideas expressed in the thesis and will give some suggestions for further work.

Chapter 2

Dropout Modelling

2.1 Introduction

In this chapter we will consider six general strategies for handling missing data : 1) Complete Case Analysis (CC), 2) Observed Data Analysis (Obs), 3) Inverse Probability Weighted Estimating Equations with Missing At Random Assumption (IPW), 4) Linear Mixed Effect models (LME), 5) Linear Increment models (LI) and 6) Last Observation Carried Forward (LOCF). We define each method and illustrate how to apply it to estimate longitudinal model parameters under dropout. In this chapter, we focus on comparing the results obtained under the dropout models mentioned in the Introduction. These models are MCAR, MAR, MNAR and SP. All of the methods are compared using simulated and real data with two timepoints or three timepoints. The most used references here are for example Diggle et al. [2007], Fitzmaurice et al. [2012] Matthews et al. [2012] and Philipson et al. [2008].

2.2 Methods of Analysis

2.2.1 Complete Case analysis (CC)

By definition, a Complete Case analysis method includes only complete cases with all measurements recorded. This approach is simply based on excluding the units that have missingness at any timepoints. The analysis involves only those individuals who have all the measurements at all the timepoints from the baseline to the endpoint and also have observed values for all relevant covariates, see Carpenter and Bartlett[1].

We can analyse any data set, and most software provides such analysis routines if we ignore all those cases whose response was not recorded for at least one of the measurements on an intended occasion. Since the remaining observations have a complete record for all the intended data, we call it Complete Case analysis. This of course results in loss of

---

[1]J. Carpenter and J. Bartlett. Simple ad-hoc methods for coping with missing data. Web-site. URL http://missingdata.lshtm.ac.uk. [Accessed: 2016-09-02].

information as the estimation is for the individuals with complete records and no effort is made to estimate the dropout mechanism. It may result in biased or unbiased results. When the observations are missing MCAR we can exclude any number of observations and the estimation procedures are unbiased. Indeed, in the no missingness situation, an estimator should be consistent, then the derived complete case analysis is consistent only if MCAR is valid, (Molenberghs et al. [2004]).

This type of analysis provides a number of advantages. A clear advantage is its easy application and it is simple to describe and therefore any software will be able to handle it since it comes with no missing data. Also, as another advantage, this type of analysis provides valid results in the case of MCAR. However, this method has its own disadvantages, excluding individuals with incomplete data may result in the loss of statistical power due to inefficient estimates. First, there is often huge loss of information. For example, suppose there are 20 measurements, with 10% of missing data on each measurement. Further, assume that, on the different measurements, the missingness is independent; then, the estimated percentage of incomplete observations is as high as 87%. This may result in huge impact on both precision and power. Even though the reduction of the number of complete cases will be less severe in the settings that have correlated missingness indicators, the loss of information will usually militate against a complete case analysis. Second, when the missingness mechanism is MAR instead of MCAR, a severe bias is possible.

2.2.2 Observed Case analysis (Obs)

Observed Case analysis uses as large as possible set of available data to estimate parameters. The approach is to exclude units with no data at all intended occasions. This approach provides an estimation procedure for all observed subjects, hence we do not use the dropout probabilities in this analysis. For example, one longitudinal study with Observed Case analysis will include all individuals who have any observed responses at any scheduled timepoint. This type of analysis is easy to implement and provides valid results in the case of MCAR. Otherwise, the analysis may again produce biased estimates. Some references call this type of analysis *available case analysis*, see for instance Gelman and Hill [2007].

2.2.3 Last Observation Carried Forward (LOCF)

The Last Observation Carried Forward method is applied in the case when longitudinal measurements are observed for each individual. The LOCF method assumes the same value for all the subsequent missing values based on the last available response. The procedure is to carry forward the last observed value or in other words to take the last observed response and substitute that value into all subsequent missing values. This al-

lows us to pretend that the data are complete since the missing values are filled in, see Carpenter and Bartlett[2]. For example, in the two timepoints case, let $Y^{R_i}=(Y_{i1}, Y_{i2})$ if $R_i=1$ and $Y^{R_i}=(Y_{i1}, Y_{i1})$ if $R_i=0$.

The LOCF method could incur problems if early dropouts occur and if the response variable has expected changes over time. It can provide biased treatment comparisons if there are different rates of dropout at different times. Diggle et al. [2007] noted that there are several weakness of the LOCF method, for example imputing fixed values at best means ignoring the random variation. One way to fix this disadvantage is hot deck imputation which works by sampling those values who dropped out later from a distribution. This distribution could be an empirical distribution or a distributional model. For more details see Rubin [1987]. The LOCF method is easy to use and has been adopted in many applications and disciplines, such as in pharmaceutical trials, despite the valid criticism it attracted. In general, LOCF is not a recommended method, (Molenberghs et al. [2004]).

2.2.4   Linear Mixed Effect (LME)

A statistical model containing fixed effects and random effects is called a mixed effect model. These models have been shown to be effective in many disciplines in the biological, physical, and social sciences. Usually a linear form is assumed.

Diggle et al. [2007] gave a definition of the response Y in the LME model which is of the form:

$$\begin{pmatrix} measured \\ response \end{pmatrix} = \begin{pmatrix} covariate \\ effects \end{pmatrix} + \begin{pmatrix} random \\ effects \end{pmatrix} + \begin{pmatrix} measurement \\ error \end{pmatrix}.$$

For example a simplified version of the Laird and Ware [1982] mixed model approach for longitudinal data would include a random effect in the intercept term in a model for responses. If Y_{ij} is the response at time j on subject i, the model is

$$Y_{ij} = \mu_{ij} + U_i + \epsilon_{ij}$$

where μ_{ij} is the marginal mean, which will usually be a linear function of covariates, ϵ_{ij} is independent Gaussian noise, and U_i is a realisation of a zero mean scalar Gaussian random variable. Since U_i has zero mean, the marginal mean of Y_{ij} remains μ_{ij} after integrating out U_{ij}. But since U_i is common to all j, we get dependence between observations on the same subject. For example if U_i is positive then all values would tend to be above the marginal mean and so on.

More complex models, with vectors of random effects, can be used to describe both more

---

[2]J. Carpenter and J. Bartlett. Simple ad-hoc methods for coping with missing data. Web-site. URL http://missingdata.lshtm.ac.uk. [Accessed: 2016-09-02].

complex correlation structures and subject specific covariate effects. A good example is when trends in time vary between subjects, in which case we might have

$$Y_{ij} = \mu_{ij} + U_{1i} + U_{2i}t_j + \epsilon_{ij}.$$

Here (U_{1i}, U_{2i}) is bivariate Gaussian. In the context of longitudinal data, some reviews of linear mixed models can be found in Cnaan et al. [1997] and Molenberghs and Verbeke [2001].

2.2.5 Linear Increments (LI)

Under this approach, the assumption is that increments in responses follow linear models and the within subject random effects have a martingale structure. This method is used for processes when we assume the parameter β to vary with time, which is the parameter in the linear predictor for the mean. However it can give estimates in the case of time constant parameters too. We will describe the rationale for this approach for the two and three treatment situations. This will be discussed in more detail in the next chapter. A simple special case is considered here.

- For two timepoints case:
  First, suppose the models for the first and second time response are respectively:
  $Y_{i1} = \mu_{i1} + U_i + \epsilon_{i1}$, where $\mu_{i1} = \beta_1 + \beta_2 x_i$, $Y_{i2} = \mu_{i2} + U_i + \epsilon_{i2}$, where $\mu_{i2} = \beta_3 + \beta_4 x_i$.
  We can rewrite as:
  $Y_{i1} = \beta_1 + \beta_2 x_i + U_i + \epsilon_{i1}$, $Y_{i2} = \beta_3 + \beta_4 x_i + U_i + \epsilon_{i2}$
  Now let $D_i = Y_{i2} - Y_{i1}$. Then we have the increment: $D_{1i} = (\beta_3 - \beta_1) + (\beta_4 - \beta_2)x_i + \epsilon_{i1}^*$, where $\epsilon_{i1}^* = \epsilon_{i2} - \epsilon_{i1}$.
  If we set $\beta_3 - \beta_1 = \gamma_1$ and $\beta_4 - \beta_2 = \gamma_2$, then $E(D_{1i}) = E[\gamma_1 + \gamma_2 x_i]$, which means we need to estimate the regression coefficients $\hat{\gamma_1}$ and $\hat{\gamma_2}$.

- For the three timepoints case, in addition to what was defined for the two timepoints case, define $D_{2i} = Y_{i3} - Y_{i2}$ The model for the third time response might be:
  $Y_{i3} = \beta_5 + \beta_6 x_i + U_i + \epsilon_{i3}$.
  Then we have the increments: $D_{1i} = \gamma_1 + \gamma_2 x_i + \epsilon_{i1}^*$, and $D_{2i} = \gamma_3 + \gamma_4 x_i + \epsilon_{i2}^*$, where $\epsilon_{i2}^* = \epsilon_{i3} - \epsilon_{i2}$, and $\gamma_3 = \beta_5 - \beta_3$, $\gamma_4 = \beta_6 - \beta_4$.
  Then $E(D_{2i}) = E[\gamma_3 + \gamma_4 x_i]$ which means we need to find the estimated regression coefficients $\hat{\gamma_1}$, $\hat{\gamma_2}$, $\hat{\gamma_3}$ and $\hat{\gamma_4}$.

Diggle et al. [2007] showed that a model for the response Y can be defined generally in terms of linear models on its increments. A full model for this approach will be discussed in the next chapter. Thereafter, Elgmati et al. [2010] extended the work of Diggle et al. [2007] within the setting of generalized estimating equations.

One of the most important assumptions of the LI approach is that the residuals ϵ_{ij}^* follow a martingale structure, in other words $E[\epsilon_{ij}^* | W_j] = 0$ where W_j denotes the history of all

observations up to time j. There are no further assumptions on the ϵ_{ij}^*, i.e. there is no assumption of Gaussian or other distribution, and no requirement for common variance between or within individuals according to Farewell [2006].

Diggle et al. [2007] state that "However, the martingale residuals do impose a condition on the mean of their distribution given their past. This single condition, of unbiased estimation of the future by the past, is sufficiently strong to be easily dismissed in many application areas though we note that this can often be overcome by suitable adjustment of the linear model".

2.2.6   Inverse Probability Weighted Estimating Equations with MAR assumption, IPW-MAR

In this method, the assumptions required are a linear model for the mean and a parametric model for dropout probability. The general idea is to construct weights for complete cases in order to reduce or remove bias. The sample is divided into subgroups. Each subgroup includes those individuals with the same baseline and similar response pattern. Rotnitzky et al. [1998] introduced a weighting method under the MAR assumption to correctly analyse a generalized estimating equation (GEE). We assume $E[Y_{ij}|x_{ij}]=x_{ij}\beta$, where $i = 1, \ldots, n$ and $j = 1, 2$ (in case we have two timepoints) or $j = 1, 2, 3$ (in case we have three timepoints) .

First, consider the two timepoints case:
The assumed model is MAR, which is defined in equation (1.5) in the Introduction, ignoring the superscripts, is: $\pi_i(\theta)=P(R = 1|Y_{i1}, Y_{i2}) = \frac{e^{\theta_0+\theta_1 Y_{i1}}}{1+e^{\theta_0+\theta_1 Y_{i1}}}$.

We fit the assumed model by logistic regression. So we choose $\hat{\theta} = \begin{pmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \end{pmatrix}$ to solve

$$\frac{1}{n} \sum \left\{ \frac{R_i}{\pi_i(\theta)} - \frac{1 - R_i}{1 - \pi_i(\theta)} \right\} \begin{pmatrix} \frac{\partial \pi_i}{\partial \theta_0} \\ \frac{\partial \pi_i}{\partial \theta_1} \end{pmatrix} = 0.$$

For complete data, we would estimate the parameter β by solving the GEE

$$\sum_{i=1}^{n} \{x_{i1}^T(Y_{i1} - x_{i1}\beta) + x_{i2}^T(Y_{i2} - x_{i2}\beta)\} = 0. \tag{2.1}$$

When data are MCAR, there is consistent estimation of β using only the observed data. As previously, R_i is the indicator of the observation at time 2, so if individual i was observed at time 2 thus $R_i=1$ and $R_i=0$ otherwise. Then the observed data GEE is

$$\sum_{i=1}^{n} \{x_{i1}^T(Y_{i1} - x_{i1}\beta) + R_i x_{i2}^T(Y_{i2} - x_{i2}\beta)\} = 0. \tag{2.2}$$

If the data are MAR an alternative GEE is

$$\sum_{i=1}^{n} \frac{R_i}{\pi_i(\theta)} \{x_{i1}^T(Y_{i1} - x_{i1}\beta) + x_{i2}^T(Y_{i2} - x_{i2}\beta)\} + (1 - \frac{R_i}{\pi_i(\theta)})\phi_1(Y_{iO}) = 0. \qquad (2.3)$$

Here, $\phi_1(Y_{iO})$ is a vector with the same dimension as β to be chosen by the analyst. This will give consistent estimates of β and is called the inverse probability weighted GEE. Determining the correct ϕ_1 is quite difficult, therefore as suggested in Robins et al. [1995] we will assume $\phi_1 = 0$.

When data are MNAR, the dropout model here is equation (1.7) which is defined in the Introduction. Recall

$$\pi_i(\theta, \theta_2) = P(R = 1|Y_{i1}, Y_{i2}) = \frac{e^{\theta_0 + \theta_1 Y_{i1} + \theta_2 Y_{i2}}}{1 + e^{\theta_0 + \theta_1 Y_{i1} + \theta_2 Y_{i2}}}.$$

Thus it is clear that $P(R_i = 1|Y_{i1}, Y_{i2})$ depends on Y_{i2} and the estimation of $\pi(\theta, \theta_2)$ is not straightforward because some of the Y_{i2} are missing and these are needed for the estimation procedure. Since the estimation of θ_2 is so difficult, a two stage sensitivity analysis is recommended by Robins et al. [1995]. These authors suggested first to estimate the dropout parameters (θ_0, θ_1) for a specified value of θ_2, by for example the solution $\hat{\theta}$, of

$$\sum_{i=1}^{n}\{1 - \frac{R_i}{\pi_i(\theta, \theta_2)}\}\pi_i(\theta) \begin{pmatrix} 1 \\ Y_{i1} \end{pmatrix} = 0 \qquad (2.4)$$

where

$$\pi_i(\theta, \theta_2) = P(R_i = 1|Y_{1i}, Y_{2i}) = \frac{e^{\theta_0 + \theta_1 Y_{1i} + \theta_2 Y_{2i}}}{1 + e^{\theta_0 + \theta_1 Y_{1i} + \theta_2 Y_{2i}}}.$$

In the second step, the estimate of β can be obtained by solving (2.3) with $\pi_i(\theta)$ replaced by $\pi_i(\hat{\theta}, \theta_2)$. The influence on inference about $\hat{\beta}$ can be monitored by repeating this procedure through a range of plausible values of θ_2.

Note that when data are MAR there is a closed form for the estimator:

$$\hat{\beta} = \left(\sum_{i=1}^{n} \frac{R_i}{\pi(\hat{\theta}, \theta_2)} \{x_{i1}^T x_{i1} + x_{i2}^T x_{i2}\}\right)^{-1} \times \left(\sum_{i=1}^{n} \frac{R_i}{\pi(\hat{\theta}, \theta_2)} \{x_{i1}^T Y_{i1} + x_{i2}^T Y_{i2}\}\right) \qquad (2.5)$$

Now, consider the three timepoints case. The observation probabilities are:

$$P(R_{i1} = 1|Y_{i1}, Y_{i2}) = \frac{\exp(\theta_0 + \theta_1 Y_{i1} + \theta_2 Y_{i2})}{1 + \exp(\theta_0 + \theta_1 Y_{i1} + \theta_2 Y_{i2})}$$

$$P(R_{i2} = 1|Y_{i2}, Y_{i3}) = \frac{\exp(\theta_0 + \theta_1 Y_{i2} + \theta_2 Y_{i3})}{1 + \exp(\theta_0 + \theta_1 Y_{i2} + \theta_2 Y_{i3})}$$

and a similar method applies as in the two timepoints case. Obviously, this can be generalised to more than three timepoints.

2.3 Simulations

We demonstrate the different estimation methods in two simulation studies, two timepoints and three timepoints examples. For each set of parameters we consider sample sizes of 250, 500, and 1000. We simulate MCAR, MAR, MNAR and SP.

2.3.1 Two timepoints simulated data

For each of $n =$250, 500, and 1000 we took 100 simulations from the model as follows. First we generated a scalar $N(0,1)$ variable x, then we generated the longitudinal means $\mu_1 = \beta_1 + \beta_2 x$, $\mu_2 = \beta_3 + \beta_4 x$. This was followed by (Y_1, Y_2) from a bivariate normal distribution with mean (μ_1, μ_2), common variance $\sigma^2 = \sigma_1^2 = \sigma_2^2$ and correlation ρ obtained through a shared parameter U. Missingness for MCAR, MAR and MNAR was generated from the logistic model:

$$\pi(\theta, \theta_2) = P(R = 1 | Y_1, Y_2) = \frac{e^{\theta_0 + \theta_1 Y_1 + \theta_2 Y_2}}{1 + e^{\theta_0 + \theta_1 Y_1 + \theta_2 Y_2}}$$

where $\theta = (\theta_0, \theta_1)$.

Missingness for SP was generated from the logistic model:

$$\pi(\theta) = P(R = 1 | Y_1, Y_2) = \frac{e^{\theta_0 + \theta_1 U}}{1 + e^{\theta_0 + \theta_1 U}}.$$

We used $\sigma^2 = 1$, $\rho = 0.5$, $\beta = (-2, -2, -1, -1)$. For missing completely at random (MCAR) we took $\theta = (1.1, 0)$ and $\theta_2 = 0$, which gave 25% missingness. Results are presented in Table 2.1. And we also took $\theta = (0, 0)$ and $\theta_2 = 0$, which gave 50% missingness, result are shown in Table 2.2. For missing at random (MAR) we used $\theta = (2.5, 0.5)$ and $\theta_2 = 0$, which gave 25% missingness, in Table 2.3, and $\theta = (1, 0.5)$ and $\theta_2 = 0$, which gave 50% missingness, presented in Table 2.4. For missing not at random (MNAR) we used $\theta = (1.5, 0)$ and $\theta_2 = 0.5$, which gave 25% missingness and we used $\theta = (0.5, 0)$ and $\theta_2 = 0.5$, which gave 50% missingness. Results are shown in Table 2.5 and 2.6 respectively. Finally, for SP we used $\theta = (1, 1)$, which gave 25% missingness and we used $\theta = (0, 0.7)$, which gave 50% missingness. Results are shown in Tables 2.7 and 2.8 respectively.

Table 2.1 shows the result at 25% missingness. All methods except LOCF seem to work as we see there is no bias. The LOCF method has bias as (β_3, β_4) differers from (β_1, β_2) meaning that carrying forward Y_1 in place of missing Y_2 is a mistaken strategy. The standard errors get smaller as the sample size increases for all the methods. Also, β_1 and β_2 generally have smaller standard errors than β_3 and β_4 at any sample size except for the CC method, where β_1, β_2, β_3 and β_4 all have about the same standard errors at each sample size. The CC and IPW methods often have higher standard error than the other methods. Note that CC and Obs give the same results for β_3 and β_4, as both are based on the available data at time 2. The methods differ at time 1, where the observed data

method makes use of the first time observations for subjects who subsequently drop out, whereas CC does not.

Table 2.2 shows the result at 50% missingness. Similar to the previous result, there is no evidence of bias except for LOCF method. The IPW results have smaller standard errors than in Table 2.1. Again CC and IPW tend to have higher standard errors than the other methods.

Table 2.3 shows results under 25% MAR missingness. For the estimates of β_3 and β_4 we see considerable upward bias under CC and Obs methods (β_3 and β_4 tends to be larger than the true values), some downward bias under LI and LOCF (β_3 and β_4 tends to be smaller than the true values), but no bias under IPW and LME, which are both consistent procedures under MAR dropout. Standard errors are typically similar to Table 2.1 for MCAR.

The reason for the upward bias under CC and Obs is that we have parametrised so that $\theta_1 = 0.5$, meaning high values of Y_1 lead to high continuation probabilities. Thus the observed data contain higher than usual values. For LI on the other hand the estimates at time 2 are based on the differences $Y_2 - Y_1$, for completers. Having higher than expected Y_1 for completers leads to lower than expected differences and negative bias. For LOCF, the Y_1 values of dropouts tend to be low, and extrapolating these to the second timepoint again leads to negative bias.

Table 2.4 has results for MAR with 50% missingness. The upward bias for CC and Obs is more severe as is the downward bias for LI and LOCF. Standard errors are higher, and again there is particular instability for IPW.

Table 2.5 shows results under 25% MNAR missingness. We see considerable upward bias for β_3 and β_4 under all methods except under LOCF, for which there is downward bias. Standard errors are typically similar to Table 2.1 for MCAR.

Table 2.6 has results for MNAR with 50% missingness. The upward bias for CC, Obs, IPW and LME is again more severe as is the downward bias for LOCF. Standard errors are higher.

Table 2.7 shows results under 25% SP missingness. All methods give consistent estimates for β_2. All methods except LOCF give consistent estimates for β_4. For β_3 we see considerable upward bias under all methods except under LOCF where there is downward bias, but no bias under LI, which is a consistent procedure under SP dropout. Standard errors are typically similar to Table 2.1 for MCAR.

Table 2.8 has results for SP with 50% missingness. The upward bias for CC, Obs, IPW

and LME is more severe as is the downward bias for LOCF as expected. Standard errors are higher.

Table 2.1: MCAR with 25% missing and two timepoints. Simulation results are based on 100 replicates for each combination. Mean estimates of coefficients β from batches of 100 simulations (standard errors in brackets).

Sample size	True	CC	Obs	IPW	LME	LI	LOCF
n=250	$\beta_1 = -2$	-1.95	-1.99	-1.96	-1.99	-1.99	-1.99
		(0.13)	(0.06)	(0.12)	(0.06)	(0.06)	(0.06)
	$\beta_2 = -2$	-1.99	-2.01	-1.99	-2.01	-2.01	-2.01
		(0.14)	(0.07)	(0.14)	(0.07)	(0.07)	(0.07)
	$\beta_3 = -1$	-0.96	-0.96	-0.96	-0.97	-0.99	-1.75
		(0.11)	(0.11)	(0.11)	(0.11)	(0.13)	(0.07)
	$\beta_4 = -1$	-0.98	-0.98	-0.98	-0.99	-1.01	-1.79
		(0.15)	(0.15)	(0.15)	(0.14)	(0.14)	(0.08)
n=500	$\beta_1 = -2$	-2.01	-2.00	-2.01	-2.00	-2.00	-2.00
		(0.09)	(0.05)	(0.08)	(0.05)	(0.05)	(0.05)
	$\beta_2 = -2$	-2.00	-2.00	-2.00	-2.00	-2.00	-2.00
		(0.10)	(0.06)	(0.10)	(0.06)	(0.06)	(0.06)
	$\beta_3 = -1$	-1.01	-1.01	-1.01	-1.01	-1.00	-1.78
		(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	(0.05)
	$\beta_4 = -1$	-0.99	-0.99	-0.99	-0.99	-0.99	-1.76
		(0.11)	(0.11)	(0.11)	(0.10)	(0.10)	(0.07)
n=1000	$\beta_1 = -2$	-2.01	-1.99	-2.01	-1.99	-1.99	-1.99
		(0.07)	(0.03)	(0.06)	(0.03)	(0.03)	(0.03)
	$\beta_2 = -2$	-2.00	-2.00	-2.00	-2.00	-2.00	-2.00
		(0.08)	(0.03)	(0.08)	(0.03)	(0.03)	(0.03)
	$\beta_3 = -1$	-0.98	-0.98	-0.98	-0.97	-0.96	-1.76
		(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.04)
	$\beta_4 = -1$	-1.00	-1.00	-1.01	-1.00	-1.00	-1.77
		(0.07)	(0.07)	(0.07)	(0.06)	(0.07)	(0.05)

Table 2.2: MCAR with 50% missing and two timepoints. Simulation results are based on 100 replicates for each combination. Mean estimates of coefficients β from batches of 100 simulations (standard errors in brackets).

Sample size	True	CC	Obs	IPW	LME	LI	LOCF
n=250	$\beta_1 = -2$	-2.01	-2.01	-2.01	-2.01	-2.01	-2.01
		(0.10)	(0.06)	(0.09)	(0.06)	(0.06)	(0.06)
	$\beta_2 = -2$	-2.01	-2.00	-2.01	-2.00	-2.00	-2.00
		(0.10)	(0.06)	(0.10)	(0.06)	(0.06)	(0.06)
	$\beta_3 = -1$	-1.01	-1.01	-1.01	-1.01	-1.01	-1.53
		(0.10)	(0.10)	(0.10)	(0.10)	(0.11)	(0.08)
	$\beta_4 = -1$	-1.00	-1.00	-1.00	-1.00	-1.00	-1.52
		(0.09)	(0.09)	(0.09)	(0.08)	(0.10)	(0.08)
n=500	$\beta_1 = -2$	-2.01	-2.00	-2.02	-2.00	-2.00	-2.00
		(0.07)	(0.05)	(0.07)	(0.05)	(0.05)	(0.05)
	$\beta_2 = -2$	-1.98	-1.99	-1.98	-1.99	-1.99	-1.99
		(0.08)	(0.05)	(0.08)	(0.05)	(0.05)	(0.05)
	$\beta_3 = -1$	-1.00	-1.00	-1.01	-1.00	-1.00	-1.53
		(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)
	$\beta_4 = -1$	-0.99	-0.99	-0.99	-0.99	-1.00	-1.54
		(0.07)	(0.07)	(0.07)	(0.06)	(0.07)	(0.06)
n=1000	$\beta_1 = -2$	-2.00	-2.00	-2.00	-2.00	-2.00	-2.00
		(0.05)	(0.03)	(0.05)	(0.03)	(0.03)	(0.03)
	$\beta_2 = -2$	-2.01	-2.00	-2.01	-2.00	-2.00	-2.00
		(0.05)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
	$\beta_3 = -1$	-0.99	-0.99	-0.99	-0.99	-1.00	-1.52
		(0.05)	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)
	$\beta_4 = -1$	-1.00	-1.00	-1.00	-1.00	-1.00	-1.52
		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)

Table 2.3: MAR with 25% missing and two timepoints. Simulation results are based on 100 replicates for each combination. Mean estimates of coefficients β from batches of 100 simulations (standard errors in brackets).

Sample size	True	CC	Obs	IPW	LME	LI	LOCF
n=250	β_1= -2	-1.91	-2.00	-2.01	-2.00	-2.00	-2.00
		(0.07)	(0.07)	(0.08)	(0.07)	(0.07)	(0.07)
	β_2=-2	-1.93	-1.99	-1.98	-1.99	-1.99	-1.99
		(0.07)	(0.06)	(0.08)	(0.06)	(0.06)	(0.06)
	β_3=-1	-0.94	-0.94	-0.99	-0.99	-1.04	-1.40
		(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)
	β_4=-1	-0.97	-0.97	-0.99	-1.00	-1.03	-1.44
		(0.07)	(0.07)	(0.08)	(0.07)	(0.06)	(0.09)
n=500	β_1= -2	-1.88	-2.00	-1.99	-2.00	-2.00	-2.00
		(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
	β_2=-2	-1.94	-2.01	-2.00	-2.01	-2.01	-2.01
		(0.06)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
	β_3=-1	-0.94	-0.94	-0.99	-0.99	-1.05	-1.41
		(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)
	β_4=-1	-0.97	-0.97	-1.01	-1.01	-1.04	-1.45
		(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)
n=1000	β_1= -2	-1.89	-2.00	-2.00	-2.00	-2.00	-2.00
		(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
	β_2=-2	-1.93	-2.00	-2.00	-2.00	-2.00	-2.00
		(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
	β_3=-1	-0.95	-0.95	-1.00	-1.00	-1.06	-1.41
		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
	β_4=-1	-0.97	-0.97	-1.00	-1.00	-1.03	-1.46
		(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)

Table 2.4: MAR with 50% missing and two timepoints. Simulation results are based on 100 replicates for each combination. Mean estimates of coefficients β from batches of 100 simulations (standard errors in brackets).

Sample size	True	CC	Obs	IPW	LME	LI	LOCF
n=250	$\beta_1 = -2$	-1.77	-1.99	-1.99	-1.99	-1.99	-1.99
		(0.09)	(0.06)	(0.09)	(0.06)	(0.06)	(0.06)
	$\beta_2 = -2$	-1.92	-2.00	-1.98	-2.00	-2.00	-2.00
		(0.09)	(0.05)	(0.13)	(0.05)	(0.05)	(0.05)
	$\beta_3 = -1$	-0.88	-0.88	-0.99	-0.99	-1.10	-1.73
		(0.10)	(0.10)	(0.11)	(0.08)	(0.08)	(0.06)
	$\beta_4 = -1$	-0.96	-0.96	-1.00	-1.00	-1.04	-1.71
		(0.10)	(0.10)	(0.14)	(0.09)	(0.10)	(0.09)
n=500	$\beta_1 = -2$	-1.78	-2.01	-2.01	-2.01	-2.01	-2.01
		(0.05)	(0.03)	(0.07)	(0.03)	(0.03)	(0.03)
	$\beta_2 = -2$	-1.91	-1.99	-2.00	-1.99	-1.99	-1.99
		(0.07)	(0.05)	(0.11)	(0.05)	(0.05)	(0.05)
	$\beta_3 = -1$	-0.89	-0.89	-1.01	-1.00	-1.12	-1.75
		(0.07)	(0.07)	(0.08)	(0.07)	(0.07)	(0.05)
	$\beta_4 = -1$	-0.95	-0.95	-1.01	-1.00	-1.04	-1.70
		(0.06)	(0.06)	(0.10)	(0.05)	(0.06)	(0.06)
n=1000	$\beta_1 = -2$	-1.76	-2.00	-2.00	-2.00	-2.00	-2.00
		(0.05)	(0.03)	(0.06)	(0.03)	(0.03)	(0.03)
	$\beta_2 = -2$	-1.91	-2.00	-2.01	-2.00	-2.00	-2.00
		(0.06)	(0.04)	(0.08)	(0.04)	(0.04)	(0.04)
	$\beta_3 = -1$	-0.87	-0.87	-0.99	-0.99	-1.11	-1.74
		(0.05)	(0.05)	(0.06)	(0.04)	(0.05)	(0.04)
	$\beta_4 = -1$	-0.95	-0.95	-0.99	-0.99	-1.04	-1.70
		(0.05)	(0.05)	(0.07)	(0.05)	(0.05)	(0.04)

Table 2.5: MNAR with 25% missing and two timepoints. Simulation results are based on 100 replicates for each combination. Mean estimates of coefficients β from batches of 100 simulations (standard errors in brackets).

Sample size	True	CC	Obs	IPW	LME	LI	LOCF
$n=250$	$\beta_1= -2$	-1.93	-1.99	-2.00	-1.99	-1.99	-1.99
		(0.07)	(0.08)	(0.07)	(0.08)	(0.08)	(0.08)
	$\beta_2=-2$	-1.97	-2.00	-1.99	-2.00	-2.00	-2.00
		(0.07)	(0.07)	(0.08)	(0.07)	(0.07)	(0.07)
	$\beta_3=-1$	-0.86	-0.86	-0.90	-0.90	-0.93	-1.33
		(0.07)	(0.07)	(0.07)	(0.07)	(0.08)	(0.09)
	$\beta_4=-1$	-0.97	-0.97	-0.98	-0.98	-1.00	-1.40
		(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.10)
$n=500$	$\beta_1= -2$	-1.93	-2.00	-2.00	-2.00	-2.00	-2.00
		(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
	$\beta_2=-2$	-1.96	-1.99	-1.98	-1.99	-1.99	-1.99
		(0.05)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
	$\beta_3=-1$	-0.86	-0.86	-0.90	-0.90	-0.93	-1.34
		(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)
	$\beta_4=-1$	-0.95	-0.95	-0.96	-0.97	-0.98	-1.39
		(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.07)
$n=1000$	$\beta_1= -2$	-1.93	-2.00	-2.00	-2.00	-2.00	-2.00
		(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	$\beta_2=-2$	-1.98	-1.99	-2.00	-1.99	-1.99	-1.99
		(0.05)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
	$\beta_3=-1$	-0.87	-0.87	-0.90	-0.90	-0.94	-1.34
		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
	$\beta_4=-1$	-0.95	-0.95	-0.96	-0.96	-0.97	-1.38
		(0.05)	(0.05)	(0.05)	(0.04)	(0.05)	(0.06)

Table 2.6: MNAR with 50% missing and two timepoints. Simulation results are based on 100 replicates for each combination. Mean estimates of coefficients β from batches of 100 simulations (standard errors in brackets).

Sample size	True	CC	Obs	IPW	LME	LI	LOCF
n=250	β_1= -2	-1.89	-1.99	-2.01	-1.99	-1.99	-1.99
		(0.09)	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)
	β_2=-2	-1.97	-2.00	-2.00	-2.00	-2.00	-2.00
		(0.08)	(0.07)	(0.08)	(0.07)	(0.07)	(0.07)
	β_3=-1	-0.77	-0.77	-0.83	-0.82	-0.88	-1.55
		(0.09)	(0.09)	(0.08)	(0.09)	(0.10)	(0.09)
	β_4=-1	-0.96	-0.96	-0.98	-0.98	-0.99	-1.62
		(0.10)	(0.10)	(0.11)	(0.10)	(0.11)	(0.10)
n=500	β_1= -2	-1.87	-2.00	-1.99	-2.00	-2.00	-2.00
		(0.06)	(0.05)	(0.07)	(0.05)	(0.05)	(0.05)
	β_2=-2	-1.95	-1.99	-1.97	-1.99	-1.99	-1.99
		(0.06)	(0.04)	(0.07)	(0.04)	(0.04)	(0.04)
	β_3=-1	-0.76	-0.76	-0.82	-0.82	-0.89	-1.56
		(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)
	β_4=-1	-0.94	-0.94	-0.96	-0.96	-0.99	-1.62
		(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.08)
n=1000	β_1= -2	-1.88	-2.00	-1.99	-2.00	-2.00	-2.00
		(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
	β_2=-2	-1.96	-1.99	-1.99	-1.99	-1.99	-1.99
		(0.06)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
	β_3=-1	-0.76	-0.76	-0.82	-0.82	-0.88	-1.55
		(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.03)
	β_4=-1	-0.94	-0.94	-0.96	-0.96	-0.97	-1.60
		(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)

Table 2.7: SP with 25% missing and two timepoints. Simulation results are based on 100 replicates for each combination. Mean estimates of coefficients β from batches of 100 simulations (standard errors in brackets).

Sample size	True	CC	Obs	IPW	LME	LI	LOCF
n=250	β_1= -2	-1.87	-2.00	-1.90	-2.00	-2.00	-2.00
		(0.07)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)
	β_2=-2	-2.01	-2.01	-2.01	-2.01	-2.01	-2.01
		(0.08)	(0.06)	(0.08)	(0.06)	(0.06)	(0.06)
	β_3=-1	-0.89	-0.89	-0.90	-0.95	-1.02	-1.30
		(0.07)	(0.07)	(0.06)	(0.06)	(0.07)	(0.08)
	β_4=-1	-1.00	-1.00	-1.00	-1.00	-1.00	-1.30
		(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)
n=500	β_1= -2	-1.87	-2.00	-1.89	-2.00	-2.00	-2.00
		(0.05)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
	β_2=-2	-1.98	-1.98	-1.99	-1.98	-1.98	-1.98
		(0.06)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
	β_3=-1	-0.87	-0.87	-0.88	-0.93	-1.00	-1.29
		(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.05)
	β_4=-1	-1.00	-1.00	-1.00	-1.00	-1.00	-1.28
		(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)
n=1000	β_1= -2	-1.87	-2.00	-1.90	-2.00	-2.00	-2.00
		(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
	β_2=-2	-1.99	-1.99	-1.99	-1.99	-1.99	-1.99
		(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
	β_3=-1	-0.88	-0.88	-0.89	-0.94	-1.01	-1.30
		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
	β_4=-1	-0.99	-0.99	-0.99	-0.99	-1.00	-1.29
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)

Table 2.8: SP with 50% missing and two timepoints. Simulation results are based on 100 replicates for each combination. Mean estimates of coefficients β from batches of 100 simulations (standard errors in brackets).

Sample size	True	CC	Obs	IPW	LME	LI	LOCF
n=250	$\beta_1 = -2$	-1.82	-1.99	-1.86	-1.99	-1.99	-1.99
		(0.08)	(0.06)	(0.08)	(0.06)	(0.06)	(0.06)
	$\beta_2 = -2$	-1.97	-1.98	-1.97	-1.98	-1.98	-1.98
		(0.09)	(0.06)	(0.09)	(0.06)	(0.06)	(0.06)
	$\beta_3 = -1$	-0.82	-0.82	-0.84	-0.91	-0.99	-1.49
		(0.08)	(0.08)	(0.08)	(0.07)	(0.08)	(0.09)
	$\beta_4 = -1$	-0.98	-0.98	-0.98	-0.99	-1.00	-1.50
		(0.08)	(0.08)	(0.09)	(0.08)	(0.09)	(0.09)
n=500	$\beta_1 = -2$	-1.82	-2.00	-1.86	-2.00	-2.00	-2.00
		(0.06)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
	$\beta_2 = -2$	-1.99	-2.00	-1.99	-2.00	-2.00	-2.00
		(0.06)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
	$\beta_3 = -1$	-0.82	-0.82	-0.84	-0.91	-1.00	-1.50
		(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)
	$\beta_4 = -1$	-0.99	-0.99	-0.99	-0.99	-1.00	-1.49
		(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.07)
n=1000	$\beta_1 = -2$	-1.83	-2.00	-1.86	-2.00	-2.00	-2.00
		(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
	$\beta_2 = -2$	-2.00	-2.00	-2.00	-2.00	-2.00	-2.00
		(0.05)	(0.03)	(0.05)	(0.03)	(0.03)	(0.03)
	$\beta_3 = -1$	-0.83	-0.83	-0.85	-0.92	-1.00	-1.50
		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
	$\beta_4 = -1$	-0.99	-0.99	-1.00	-0.99	-0.99	-1.50
		(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)

2.3.2   Three timepoints simulated data

In this simulation scenario the measurements are scheduled at times $t = 1, 2, 3$. A logistic model with linear predictors is used for the probabilities of drop out between times t and $t+1$. For each of $n =$250, 500, and 1000 we took 100 simulations from the model as follows:

First we generated a scalar $N(0, 1)$ variable x, then we generated the longitudinal means $\mu_1 = \beta_1 + \beta_2 x$, $\mu_2 = \beta_3 + \beta_4 x$ and $\mu_3 = \beta_5 + \beta_6 x$. This was followed by (Y_1, Y_2, Y_3) from a multivariate normal distribution with mean (μ_1, μ_2, μ_3), common variance σ^2 and correlation ρ again obtained by a shared random effect U. Missingness was generated from the logistic model:

$$\pi(\theta, \theta_2) = P(R_{i1} = 1 | Y_{i1}, Y_{i2}) = \frac{\exp(\theta_0 + \theta_1 Y_{i1} + \theta_2 Y_{i2})}{1 + \exp(\theta_0 + \theta_1 Y_{i1} + \theta_2 Y_{i2})}$$

$$\pi(\theta, \theta_2) = P(R_{i2} = 1 | Y_{i2}, Y_{i3}) = \frac{\exp(\theta_0 + \theta_1 Y_{i2} + \theta_2 Y_{i3})}{1 + \exp(\theta_0 + \theta_1 Y_{i2} + \theta_2 Y_{i3})}$$

Note that we use the same parameters θ and θ_2 at each dropout time, for simplicity. We used $\sigma^2 = 1$, $\rho = 0.5$, $\beta = (1, 1, 1, 1, 1, 1)$. For missing completely at random (MCAR) we took $\theta = (1.85, 0)$ and $\theta_2 = 0$ which gave around 25% missingness and $\theta = (0.9, 0)$ and $\theta_2 = 0$ which gave around 50% missingness. For missing at random (MAR) we took $\theta = (2.6, -0.5)$ and $\theta_2 = 0$ which gave around 25% missingness and $\theta = (1.45, -0.5)$ and $\theta_2 = 0$ which gave around 50% missingness. For missing not at random (MNAR) we took $\theta = (2.6, 0)$ and $\theta_2 = -0.5$ which gave around 25% missingness and $\theta = (1.45, 0)$ and $\theta_2 = -0.5$ which gave around 50% missingness. For shared parameter (SP) we took $\theta = (2.31, -1)$ which gave around 25% missingness and $\theta = (0.9, -1)$ which gave around 50% missingness. Results are presented in Tables 2.9, 2.10, 2.11 and 2.12.

Tables 2.9 and 2.10 show results for 25% and 50% missingness under MCAR. None of the methods show any bias this time including LOCF. The reason LOCF works well in this example is that (β_3, β_4) is the same as (β_1, β_2), so that Y_1 is representative of missing Y_2. Standard errors decrease with sample size and are broadly similar for the six methods, except for the CC and IPW at 50% missingness, which lead to more variable estimates. In addition, in all tables, note that at time 3, CC and Obs give the same estimates for β_5 and β_6. The reason is that the CC are subjects who completed to the final timepoint (time 3 in this example). So at that time the observed data are the complete case people. At earlier times the observed data are CC plus subjects who will dropout before the end.

Table 2.11 shows that at 25% missingness under MAR model, the LME and IPW methods work as expected, since MAR is designed for them, and even at the higher missingness percentage of 50%, we have similar results. As in Tables 2.3 and 2.4, CC and Obs provide downward bias, and LI and LOCF upward bias at times after the first. The IPW method

is unstable at the lower sample sizes with 50% missingness.

Table 2.13 shows that at 25% missingness under MNAR model, no methods work as expected and at the higher missingness percentage of 50%, we have similar results. We see considerable downward bias for β_3, β_4, β_5 and β_6 under all methods. Standard errors are typically similar to Table 2.9 for MCAR.

Table 2.15 shows that at 25% missingness under SP model, the LI method works as expected, and even at the higher missingness percentage of 50%, we have similar results. As in Tables 2.7 and 2.8, CC and Obs provide downward bias, and LOCF upward bias at times after the first. The LME method has downward bias for the intercept terms β_3 and β_5. Standard errors are typically similar to Table 2.9 for MCAR.

Table 2.9: MCAR with 25% missing and three timepoints. Mean estimates of coefficients β from batches of 100 simulations (standard errors in brackets).

Sample size	True	CC	Obs	IPW	LME	LI	LOCF
n=250	β_1=1	0.99	0.99	1.00	0.99	0.99	0.99
		(0.13)	(0.12)	(0.13)	(0.12)	(0.12)	(0.12)
	β_2=1	1.02	1.01	1.02	1.01	1.01	1.01
		(0.13)	(0.10)	(0.13)	(0.10)	(0.10)	(0.10)
	β_3=1	1.00	1.00	1.01	1.00	1.01	1.01
		(0.13)	(0.11)	(0.12)	(0.11)	(0.11)	(0.11)
	β_4=1	1.01	1.01	1.01	1.00	1.00	1.00
		(0.13)	(0.11)	(0.13)	(0.11)	(0.11)	(0.10)
	β_5=1	0.99	0.99	1.00	1.00	1.00	1.00
		(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.12)
	β_6=1	1.00	1.00	1.01	1.00	0.99	1.00
		(0.12)	(0.12)	(0.12)	(0.11)	(0.11)	(0.10)
n=500	β_1=1	0.98	0.99	0.98	0.99	0.99	0.99
		(0.09)	(0.07)	(0.08)	(0.07)	(0.07)	(0.07)
	β_2=1	0.98	0.99	0.98	0.99	0.99	0.99
		(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
	β_3=1	0.99	0.99	0.99	0.99	0.99	0.99
		(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
	β_4=1	0.98	0.98	0.98	0.98	0.98	0.99
		(0.09)	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)
	β_5=1	0.98	0.98	0.98	0.98	0.99	0.99
		(0.08)	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)
	β_6=1	0.99	0.99	0.99	0.99	0.99	0.99
		(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.08)
n=1000	β_1=1	1.00	1.01	1.00	1.01	1.01	1.01
		(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
	β_2=1	0.99	0.99	1.00	0.99	0.99	0.99
		(0.06)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
	β_3=1	0.99	1.00	1.00	1.00	1.00	1.00
		(0.07)	(0.06)	(0.07)	(0.06)	(0.07)	(0.06)
	β_4=1	0.99	0.99	0.99	0.99	0.99	0.99
		(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
	β_5=1	0.99	0.99	0.99	0.99	0.99	0.99
		(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)
	β_6=1	0.99	0.99	0.99	0.99	0.99	0.99
		(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)

Table 2.10: MCAR with 50% missing and three timepoints. Simulation results are based on 100 replicates for each combination (standard errors in brackets).

Sample size	True	CC	Obs	IPW	LME	LI	LOCF
n=250	β_1= 1	0.98	0.99	0.99	0.99	0.99	0.99
		(0.16)	(0.11)	(0.14)	(0.11)	(0.11)	(0.11)
	β_2=1	1.00	0.99	1.00	0.99	0.99	0.99
		(0.15)	(0.11)	(0.15)	(0.11)	(0.11)	(0.11)
	β_3=1	0.98	0.99	0.99	0.99	0.99	0.99
		(0.17)	(0.16)	(0.15)	(0.14)	(0.14)	(0.13)
	β_4=1	1.00	0.99	1.00	0.99	0.99	0.99
		(0.14)	(0.11)	(0.14)	(0.10)	(0.11)	(0.10)
	β_5=1	0.97	0.97	0.98	0.98	0.98	0.98
		(0.17)	(0.17)	(0.15)	(0.14)	(0.14)	(0.12)
	β_6=1	1.02	1.02	1.03	1.02	1.01	1.00
		(0.17)	(0.17)	(0.17)	(0.15)	(0.16)	(0.12)
n=500	β_1=1	0.99	0.99	1.00	0.99	0.99	0.99
		(0.12)	(0.07)	(0.09)	(0.07)	(0.07)	(0.07)
	β_2=1	1.01	1.01	1.01	1.01	1.01	1.01
		(0.11)	(0.08)	(0.11)	(0.08)	(0.08)	(0.08)
	β_3=1	1.00	0.99	1.00	0.99	0.99	0.99
		(0.10)	(0.09)	(0.09)	(0.08)	(0.08)	(0.07)
	β_4=1	1.02	1.02	1.02	1.02	1.02	1.01
		(0.11)	(0.10)	(0.11)	(0.09)	(0.09)	(0.08)
	β_5=1	0.99	0.99	1.00	0.99	0.99	0.99
		(0.11)	(0.11)	(0.10)	(0.09)	(0.09)	(0.07)
	β_6=1	1.02	1.02	1.02	1.02	1.02	1.02
		(0.12)	(0.12)	(0.12)	(0.10)	(0.11)	(0.08)
n=1000	β_1=1	1.00	1.01	1.00	1.01	1.01	1.01
		(0.07)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
	β_2=1	1.01	1.01	1.01	1.01	1.01	1.01
		(0.06)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
	β_3=1	1.01	1.00	1.01	1.01	1.01	1.01
		(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)
	β_4=1	0.99	1.00	0.99	0.99	0.99	1.00
		(0.07)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)
	β_5=1	1.01	1.01	1.00	1.01	1.00	1.00
		(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.05)
	β_6=1	1.01	1.01	1.01	1.01	1.01	1.00
		(0.08)	(0.08)	(0.08)	(0.07)	(0.08)	(0.06)

Table 2.11: MAR with 25% missing and three timepoints. Simulation results are based on 100 replicates for each combination (standard errors in brackets).

Sample size	True	CC	Obs	IPW	LME	LI	LOCF
n=250	β_1=1	0.68	0.98	0.97	0.98	0.98	0.98
		(0.13)	(0.11)	(0.13)	(0.11)	(0.11)	(0.11)
	β_2=1	0.90	1.00	0.99	1.00	1.00	1.00
		(0.12)	(0.11)	(0.14)	(0.11)	(0.11)	(0.11)
	β_3=1	0.68	0.86	0.98	0.98	1.05	1.04
		(0.12)	(0.11)	(0.13)	(0.11)	(0.11)	(0.11)
	β_4=1	0.90	0.96	1.00	1.00	1.03	1.02
		(0.12)	(0.12)	(0.14)	(0.11)	(0.12)	(0.11)
	β_5=1	0.74	0.74	0.98	0.99	1.11	1.08
		(0.14)	(0.14)	(0.14)	(0.13)	(0.14)	(0.12)
	β_6=1	0.91	0.91	0.98	0.99	1.03	1.02
		(0.12)	(0.12)	(0.13)	(0.12)	(0.13)	(0.11)
n=500	β_1=1	0.69	1.00	1.00	1.00	1.00	1.00
		(0.10)	(0.08)	(0.10)	(0.08)	(0.08)	(0.08)
	β_2=1	0.90	1.00	1.00	1.00	1.00	1.00
		(0.09)	(0.08)	(0.11)	(0.08)	(0.08)	(0.08)
	β_3=1	0.68	0.86	0.99	0.99	1.06	1.05
		(0.09)	(0.08)	(0.09)	(0.07)	(0.08)	(0.07)
	β_4=1	0.91	0.96	1.02	1.01	1.03	1.03
		(0.09)	(0.08)	(0.11)	(0.08)	(0.08)	(0.08)
	β_5=1	0.75	0.75	1.00	1.00	1.12	1.10
		(0.10)	(0.10)	(0.10)	(0.09)	(0.09)	(0.08)
	β_6=1	0.92	0.92	1.01	1.00	1.04	1.03
		(0.09)	(0.09)	(0.11)	(0.09)	(0.09)	(0.08)
n=1000	β_1=1	0.70	1.01	1.00	1.01	1.01	1.01
		(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)
	β_2=1	0.91	1.01	1.01	1.01	1.01	1.01
		(0.06)	(0.06)	(0.08)	(0.06)	(0.06)	(0.06)
	β_3=1	0.69	0.87	0.99	1.00	1.07	1.06
		(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)
	β_4=1	0.90	0.96	1.00	1.00	1.03	1.02
		(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.05)
	β_5=1	0.75	0.75	0.99	1.00	1.13	1.10
		(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
	β_6=1	0.91	0.91	0.99	0.99	1.04	1.02
		(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)

Table 2.12: MAR with 50% missing and three timepoints. Simulation results are based on 100 replicates for each combination (standard errors in brackets).

Sample size	True	CC	Obs	IPW	LME	LI	LOCF
n=250	β_1= 1	0.37	0.99	1.00	0.99	0.99	0.99
		(0.16)	(0.12)	(0.23)	(0.12)	(0.12)	(0.12)
	β_2=1	0.87	1.02	1.06	1.02	1.02	1.02
		(0.17)	(0.11)	(0.30)	(0.11)	(0.11)	(0.11)
	β_3=1	0.37	0.73	1.00	1.00	1.13	1.09
		(0.16)	(0.15)	(0.22)	(0.15)	(0.15)	(0.13)
	β_4=1	0.85	0.93	1.02	1.00	1.04	1.02
		(0.17)	(0.13)	(0.29)	(0.12)	(0.12)	(0.10)
	β_5=1	0.51	0.51	1.00	1.01	1.27	1.15
		(0.15)	(0.15)	(0.20)	(0.14)	(0.15)	(0.12)
	β_6=1	0.90	0.90	1.04	1.02	1.09	1.03
		(0.16)	(0.16)	(0.29)	(0.13)	(0.15)	(0.10)
n=500	β_1=1	0.36	1.00	0.98	1.00	1.00	1.00
		(0.10)	(0.08)	(0.15)	(0.08)	(0.08)	(0.08)
	β_2=1	0.85	1.01	1.00	1.01	1.01	1.01
		(0.12)	(0.09)	(0.16)	(0.09)	(0.09)	(0.09)
	β_3=1	0.36	0.73	0.99	1.00	1.14	1.09
		(0.10)	(0.09)	(0.16)	(0.08)	(0.08)	(0.07)
	β_4=1	0.84	0.94	1.00	1.01	1.05	1.03
		(0.12)	(0.11)	(0.19)	(0.10)	(0.10)	(0.09)
	β_5=1	0.48	0.48	0.97	0.99	1.26	1.15
		(0.11)	(0.11)	(0.15)	(0.09)	(0.09)	(0.07)
	β_6=1	0.87	0.87	1.00	1.01	1.08	1.03
		(0.11)	(0.11)	(0.17)	(0.10)	(0.12)	(0.09)
n=1000	β_1=1	0.38	1.00	1.00	1.00	1.00	1.00
		(0.07)	(0.05)	(0.10)	(0.05)	(0.05)	(0.05)
	β_2=1	0.85	1.00	1.00	1.00	1.00	1.00
		(0.07)	(0.05)	(0.11)	(0.05)	(0.05)	(0.05)
	β_3=1	0.37	0.73	1.00	1.00	1.13	1.09
		(0.07)	(0.06)	(0.11)	(0.06)	(0.06)	(0.05)
	β_4=1	0.84	0.93	1.00	1.00	1.04	1.01
		(0.08)	(0.07)	(0.11)	(0.06)	(0.06)	(0.05)
	β_5=1	0.50	0.50	1.00	1.00	1.26	1.15
		(0.08)	(0.08)	(0.10)	(0.07)	(0.07)	(0.05)
	β_6=1	0.87	0.87	1.00	1.00	1.07	1.02
		(0.07)	(0.07)	(0.12)	(0.07)	(0.08)	(0.06)

Table 2.13: MNAR with 25% missing and three timepoints. Simulation results are based on 100 replicates for each combination (standard errors in brackets).

Sample size	True	CC	Obs	IPW	LME	LI	LOCF
n=250	β_1= 1	0.75	1.00	0.98	1.00	1.00	1.00
		(0.13)	(0.11)	(0.12)	(0.11)	(0.11)	(0.11)
	β_2=1	0.91	1.00	0.97	1.00	1.00	1.00
		(0.12)	(0.11)	(0.13)	(0.11)	(0.11)	(0.11)
	β_3=1	0.70	0.81	0.91	0.89	0.94	0.95
		(0.13)	(0.12)	(0.12)	(0.11)	(0.12)	(0.11)
	β_4=1	0.90	0.94	0.95	0.97	0.98	0.99
		(0.12)	(0.12)	(0.13)	(0.11)	(0.12)	(0.11)
	β_5=1	0.69	0.69	0.86	0.87	0.94	0.95
		(0.13)	(0.13)	(0.12)	(0.12)	(0.12)	(0.11)
	β_6=1	0.90	0.90	0.93	0.96	0.98	0.99
		(0.12)	(0.12)	(0.13)	(0.12)	(0.12)	(0.11)
n=500	β_1= 1	0.76	1.00	0.98	1.00	1.00	1.00
		(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
	β_2=1	0.92	1.00	0.97	1.00	1.00	1.00
		(0.09)	(0.08)	(0.10)	(0.08)	(0.08)	(0.08)
	β_3=1	0.69	0.81	0.91	0.89	0.94	0.95
		(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.07)
	β_4=1	0.90	0.94	0.95	0.96	0.98	0.99
		(0.09)	(0.08)	(0.09)	(0.08)	(0.08)	(0.07)
	β_5=1	0.69	0.69	0.86	0.87	0.93	0.95
		(0.09)	(0.09)	(0.08)	(0.08)	(0.08)	(0.07)
	β_6=1	0.90	0.90	0.93	0.96	0.98	0.99
		(0.09)	(0.09)	(0.09)	(0.08)	(0.08)	(0.07)
n=1000	β_1= 1	0.75	1.00	0.98	1.00	1.00	1.00
		(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
	β_2=1	0.91	1.00	0.97	1.00	1.00	1.00
		(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)
	β_3=1	0.69	0.81	0.91	0.89	0.94	0.95
		(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)
	β_4=1	0.90	0.93	0.95	0.96	0.98	0.99
		(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)
	β_5=1	0.69	0.69	0.86	0.87	0.94	0.95
		(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)
	β_6=1	0.90	0.90	0.94	0.96	0.98	0.99
		(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.05)

Table 2.14: MNAR with 50% missing and three timepoints. Simulation results are based on 100 replicates for each combination (standard errors in brackets).

Sample size	True	CC	Obs	IPW	LME	LI	LOCF
$n=250$	$\beta_1= 1$	0.50	1.01	0.94	1.01	1.01	1.01
		(0.15)	(0.11)	(0.16)	(0.11)	(0.11)	(0.11)
	$\beta_2=1$	0.88	1.01	0.95	1.01	1.01	1.01
		(0.15)	(0.11)	(0.19)	(0.11)	(0.11)	(0.11)
	$\beta_3=1$	0.37	0.59	0.79	0.77	0.87	0.91
		(0.14)	(0.12)	(0.16)	(0.12)	(0.12)	(0.11)
	$\beta_4=1$	0.84	0.89	0.91	0.94	0.97	0.99
		(0.15)	(0.13)	(0.18)	(0.12)	(0.12)	(0.11)
	$\beta_5=1$	0.36	0.36	0.69	0.72	0.86	0.91
		(0.15)	(0.15)	(0.16)	(0.13)	(0.14)	(0.11)
	$\beta_6=1$	0.84	0.84	0.90	0.94	0.97	1.00
		(0.15)	(0.15)	(0.18)	(0.13)	(0.14)	(0.11)
$n=500$	$\beta_1= 1$	0.50	1.00	0.95	1.00	1.00	1.00
		(0.11)	(0.08)	(0.11)	(0.08)	(0.08)	(0.08)
	$\beta_2=1$	0.87	1.00	0.95	1.00	1.00	1.00
		(0.11)	(0.08)	(0.14)	(0.08)	(0.08)	(0.08)
	$\beta_3=1$	0.37	0.59	0.80	0.77	0.86	0.91
		(0.11)	(0.09)	(0.11)	(0.08)	(0.09)	(0.08)
	$\beta_4=1$	0.84	0.89	0.90	0.94	0.96	0.99
		(0.11)	(0.09)	(0.13)	(0.09)	(0.09)	(0.08)
	$\beta_5=1$	0.37	0.37	0.70	0.73	0.86	0.91
		(0.10)	(0.10)	(0.11)	(0.09)	(0.10)	(0.08)
	$\beta_6=1$	0.84	0.84	0.89	0.93	0.96	0.99
		(0.11)	(0.11)	(0.12)	(0.10)	(0.10)	(0.08)
$n=1000$	$\beta_1= 1$	0.49	1.00	0.95	1.00	1.00	1.00
		(0.08)	(0.05)	(0.08)	(0.05)	(0.05)	(0.05)
	$\beta_2=1$	0.87	1.00	0.95	1.00	1.00	1.00
		(0.08)	(0.05)	(0.09)	(0.05)	(0.05)	(0.05)
	$\beta_3=1$	0.37	0.59	0.80	0.77	0.86	0.91
		(0.07)	(0.06)	(0.07)	(0.06)	(0.06)	(0.05)
	$\beta_4=1$	0.84	0.89	0.91	0.94	0.96	0.99
		(0.08)	(0.06)	(0.09)	(0.06)	(0.06)	(0.05)
	$\beta_5=1$	0.37	0.37	0.70	0.73	0.87	0.91
		(0.08)	(0.08)	(0.08)	(0.07)	(0.07)	(0.05)
	$\beta_6=1$	0.84	0.84	0.89	0.93	0.97	0.99
		(0.07)	(0.07)	(0.09)	(0.07)	(0.07)	(0.06)

Table 2.15: SP with 25% missing and three timepoints. Simulation results are based on 100 replicates for each combination (standard errors in brackets).

Sample size	True	CC	Obs	IPW	LME	LI	LOCF
n=250	β_1= 1	0.60	1.00	0.81	1.00	1.00	1.00
		(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
	β_2=1	1.01	1.00	1.09	1.00	1.00	1.00
		(0.12)	(0.11)	(0.13)	(0.11)	(0.11)	(0.11)
	β_3=1	0.60	0.76	0.81	0.91	1.00	1.00
		(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
	β_4=1	1.00	1.00	1.08	1.00	1.00	1.00
		(0.12)	(0.12)	(0.13)	(0.11)	(0.12)	(0.11)
	β_5=1	0.59	0.59	0.76	0.88	0.99	1.00
		(0.11)	(0.11)	(0.12)	(0.12)	(0.12)	(0.11)
	β_6=1	1.00	1.00	1.06	1.00	1.00	1.00
		(0.12)	(0.12)	(0.13)	(0.12)	(0.12)	(0.11)
n=500	β_1= 1	0.60	1.00	0.82	1.00	1.00	1.00
		(0.08)	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)
	β_2=1	1.00	1.00	1.08	1.00	1.00	1.00
		(0.09)	(0.08)	(0.10)	(0.08)	(0.08)	(0.08)
	β_3=1	0.60	0.76	0.82	0.92	1.00	1.00
		(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
	β_4=1	1.00	1.00	1.09	1.00	1.00	1.00
		(0.08)	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)
	β_5=1	0.60	0.60	0.76	0.88	1.00	1.00
		(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	(0.08)
	β_6=1	1.01	1.01	1.07	1.00	1.00	1.00
		(0.08)	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)
n=1000	β_1= 1	0.60	1.00	0.82	1.00	1.00	1.00
		(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
	β_2=1	1.00	1.00	1.09	1.00	1.00	1.00
		(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)
	β_3=1	0.60	0.76	0.82	0.92	1.00	1.00
		(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
	β_4=1	1.00	1.00	1.09	1.00	1.00	1.00
		(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)
	β_5=1	0.60	0.60	0.76	0.88	1.00	1.00
		(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
	β_6=1	1.00	1.00	1.07	1.00	1.00	1.00
		(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)

Table 2.16: SP with 50% missing and three timepoints. Simulation results are based on 100 replicates for each combination (standard errors in brackets).

Sample size	True	CC	Obs	IPW	LME	LI	LOCF
n=250	β_1= 1	0.19	1.00	0.60	1.00	1.00	1.00
		(0.14)	(0.11)	(0.15)	(0.11)	(0.11)	(0.11)
	β_2=1	1.00	1.00	1.12	1.00	1.00	1.00
		(0.14)	(0.11)	(0.18)	(0.11)	(0.11)	(0.11)
	β_3=1	0.19	0.50	0.60	0.82	1.00	1.00
		(0.14)	(0.12)	(0.15)	(0.12)	(0.13)	(0.11)
	β_4=1	1.00	1.00	1.11	1.00	1.00	1.00
		(0.14)	(0.12)	(0.18)	(0.12)	(0.13)	(0.11)
	β_5=1	0.19	0.19	0.49	0.75	1.00	1.00
		(0.13)	(0.13)	(0.14)	(0.13)	(0.14)	(0.11)
	β_6=1	1.00	1.00	1.08	1.00	1.00	1.00
		(0.14)	(0.14)	(0.17)	(0.13)	(0.15)	(0.11)
n=500	β_1= 1	0.19	1.00	0.61	1.00	1.00	1.00
		(0.09)	(0.08)	(0.11)	(0.08)	(0.08)	(0.08)
	β_2=1	1.00	1.00	1.12	1.00	1.00	1.00
		(0.10)	(0.08)	(0.13)	(0.08)	(0.08)	(0.08)
	β_3=1	0.19	0.50	0.61	0.82	1.00	1.00
		(0.09)	(0.09)	(0.10)	(0.09)	(0.09)	(0.08)
	β_4=1	1.01	1.01	1.12	1.00	1.00	1.00
		(0.10)	(0.09)	(0.14)	(0.09)	(0.09)	(0.08)
	β_5=1	0.19	0.19	0.49	0.75	1.00	1.00
		(0.09)	(0.09)	(0.10)	(0.09)	(0.10)	(0.08)
	β_6=1	1.01	1.01	1.09	1.00	1.00	1.00
		(0.10)	(0.10)	(0.12)	(0.10)	(0.10)	(0.08)
n=1000	β_1= 1	0.19	1.00	0.61	1.00	1.00	1.00
		(0.07)	(0.05)	(0.08)	(0.05)	(0.05)	(0.05)
	β_2=1	1.00	1.00	1.13	1.00	1.00	1.00
		(0.07)	(0.06)	(0.10)	(0.06)	(0.06)	(0.06)
	β_3=1	0.19	0.50	0.61	0.82	1.00	1.00
		(0.07)	(0.06)	(0.08)	(0.06)	(0.06)	(0.05)
	β_4=1	1.00	1.00	1.12	1.00	1.00	1.00
		(0.07)	(0.06)	(0.10)	(0.06)	(0.06)	(0.06)
	β_5=1	0.19	0.19	0.49	0.75	1.00	1.00
		(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)
	β_6=1	1.00	1.00	1.09	1.00	1.00	1.00
		(0.07)	(0.07)	(0.09)	(0.06)	(0.07)	(0.06)

2.4   Applications

2.4.1   Two timepoints data

We consider data from a clinical trial with two treatments and two measurement times as introduced and analysed by Matthews et al. [2012]. The data are real but not public. The original trial was randomized and observations were taken pre-randomization and at two times post-randomization. In their work only the post-randomization observations were considered. Matthews et al. [2012] state "For confidentiality reasons we have scaled all observed responses by the mean and standard deviation of the responses at time 1 and we will refer simply to treatment and response without describing what they are". The covariates in this data are only treatment type and time. The parameter vector is $(\beta_1, \beta_2, \beta_3, \beta_4)$, ignoring the time by treatment interaction. There are 422 subjects.

At time 1, all those subjects provided a response, but 24.4% dropped out by time 2. There are 212 subjects receiving Treatment A, but only 126 provided a response at time 2 and the other 86 dropped out hence the missingness percentage is 40.6%. The dropout reason is not known. For Treatment B, there are 210 subjects, of which 193 subjects continued to time 2 and hence there are 17 that did not and this gave around 8.1% missingness. Figure 2.1 presents a summary of the data. The left plot shows the mean values at times 1 and 2 of all observed data, split into the two treatment groups. The right plot shows the means for subjects who completed the trial. Means at time 0 (with no dropout) are not shown, and both groups have the same values. The impression given by the left plot in that there is a difference between groups at time 1 but this disappears by time 2. The right plot shows a different story: There is no difference between treatment groups for those who complete the trial.

Table 2.17 summarizes the estimates of the coefficients corresponding to the six methods. The standard errors are shown in Table 2.18 which were calculated using the bootstrap based on 100 samples. Note β_1 and β_2 generally have smaller standard errors than β_3 and β_4 except for the CC method, where β_1, β_2, β_3 and β_4 all have about the same standard errors. The IPW method has higher standard error than the other methods. Note that CC and Obs give the same results for β_3 and β_4 as mentioned before in the simulation study in Section 2.3.

Figures 2.2 and 2.3 show estimated treatment effects and nominal 95% confidence intervals for the different fitting methods. Figure 2.2 shows estimated coefficients β_1 and β_2 for time 1. Looking at the left plot, for β_1, the complete case value is significantly lower than the others, which might have been expected from Figure 2.1. This is consistent with the previous simulation results. Turning to Figure 2.3 and time 2, there is high variability for IPW again. The complete case and observed data methods has the least evidence of a treatment effect, while LOCF has the most. Complete case, observed data and IPW give

very similar estimates in both plots with LME, LI and LOCF progressively moving away. Note that only one of the methods -IPW- requires the dropout process to be modelled. For this we assume a logistic model $\pi_i(\theta) = \text{expit}(\theta_0 + \theta_1 Y_{1i})$ as in equation (1.5). For reference, the estimates (and standard errors) are $\hat{\theta}_0 = 1.60\,(0.16)$ and $\hat{\theta}_1 = -1.66\,(0.18)$.

Figure 2.1: Two timepoints data. Left plot: Means of observed data. Right plot: Mean for subjects who completed the trial.
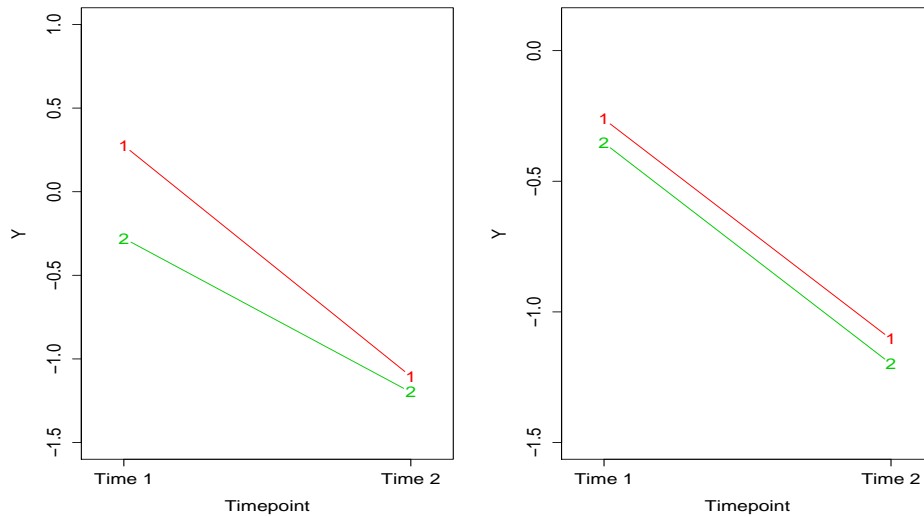
Table 2.17: Two timepoints data analysis: Estimates of coefficients β for different fitting methods.

β	CC	Obs	IPW	LME	LI	LOCF
β_1	-0.27	0.27	0.15	0.27	0.27	0.27
β_2	-0.08	-0.55	-0.18	-0.55	-0.55	-0.55
β_3	-1.10	-1.10	-0.92	-0.74	-0.56	-0.23
β_4	-0.10	-0.10	-0.08	-0.41	-0.56	-0.83

Table 2.18: Standard errors of estimates of coefficients β for different fitting methods.

s.e(β)	CC	Obs	IPW	LME	LI	LOCF
s.e(β_1)	0.07	0.07	0.13	0.07	0.07	0.07
s.e(β_2)	0.08	0.09	0.26	0.09	0.09	0.09
s.e(β_3)	0.07	0.07	0.12	0.08	0.08	0.09
s.e(β_4)	0.09	0.09	0.19	0.10	0.11	0.13

Figure 2.2: Two timepoints example: Estimated coefficients β_1 and β_2 and nominal 95% confidence intervals.
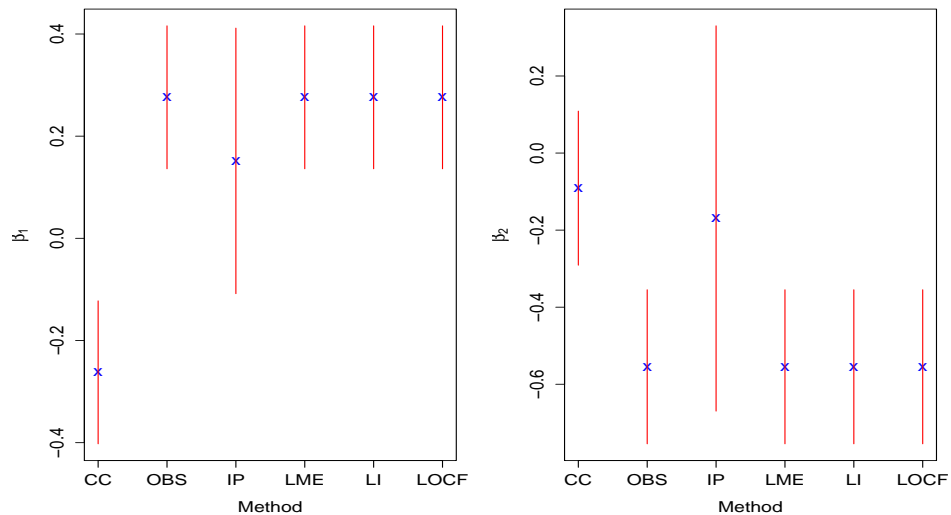
Figure 2.3: Two timepoints example: Estimated coefficients β_3 and β_4 and nominal 95% confidence intervals.
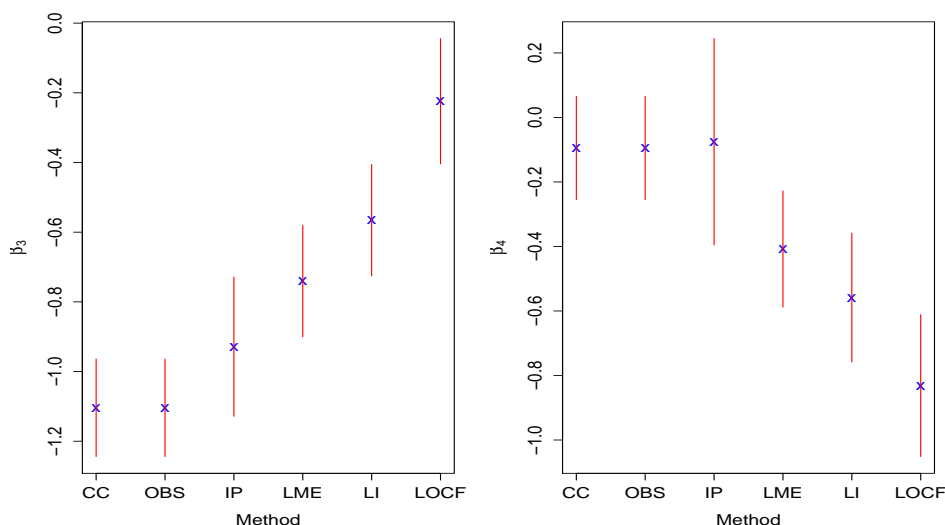
2.4.2 Three timepoints: Schizophrenia data

In this section we present a comparison of the several methods in the analysis of data from a trial into the treatment of schizophrenia. The trial compared three treatments: A standard therapy using Haloperidol (Treatment 1), a placebo (Treatment 2), and an experimental therapy using Risperidone (Treatment 3). These data are described in Henderson et al. [2000], the response is PANSS (positive and negative symptom score, a measure of mental health, with high values being bad). Values ranged from 30 to 210. Typically we expect a schizophrenia patient in this clinical trial to have score around 90.

There are in total 518 patients in this study, of which 249 dropped out. Henderson et al. [2000] stated "The goal of the study was to compare the three treatments with respect to their ability to improve (reduce) the mean PANSS score. The patients were observed at baseline ($t = 1$) and thereafter at weeks 1, 2, 4, 6 and 8 ($t=2, 3, 4, 5, 6$) of the study". The only covariates used here are treatment groups. For this work we will concentrate on three scheduled measurement times only: At the baseline, week 4, and week 8 i.e. at times $t=(1,4,6)$.

In Treatment 1 there were 85 subjects, but only 41 subjects provided a response at time 3 and the other 44 subjects dropped out which is equivalent to a 51.8% missing rate. There were 88 subjects receiving Treatment 2, but only 29 subjects provided a response at time 3 and the other 59 subjects dropped out which is equivalent to 67% missing. Of 345 subjects receiving Treatment 3, only 199 provided a response at time 3 and the other 146 dropped out which is equivalent to 42% missing. Hence dropout is the highest in treatment group 2 and the lowest in group 3.

Figure 2.4 in the left side, shows the observed mean response as a function of time within each treatment group, i.e. each average is over those patients who have not yet dropped out. All three groups have a decreasing mean response, perhaps at a slower rate towards the third timepoint. The overall reduction in mean response within each treatment group is very roughly from between 90 and 95 to around 70 and 75. This appears close to the criterion for clinical improvement, which was stated in advance of the trial, to be a reduction of 20% in the mean PANSS scores. The decrease in group 2 was smaller overall. However, at each timepoint these observed means are, necessarily, calculated only from those subjects who have not yet dropped out. Figure 2.4 in the right side, shows the behavior of complete cases: Large differences between groups are evident, with the highest decrease in the groups 1 and 3 and the lowest in the group 2. Both graphs shows the slope is dropping more in group 1 and 3, while for group 2 there is more stability (less variability).

Figure 2.4: Schizophrenia data: This figure shows the mean values of observed cases in the left side and the complete cases in the right side.
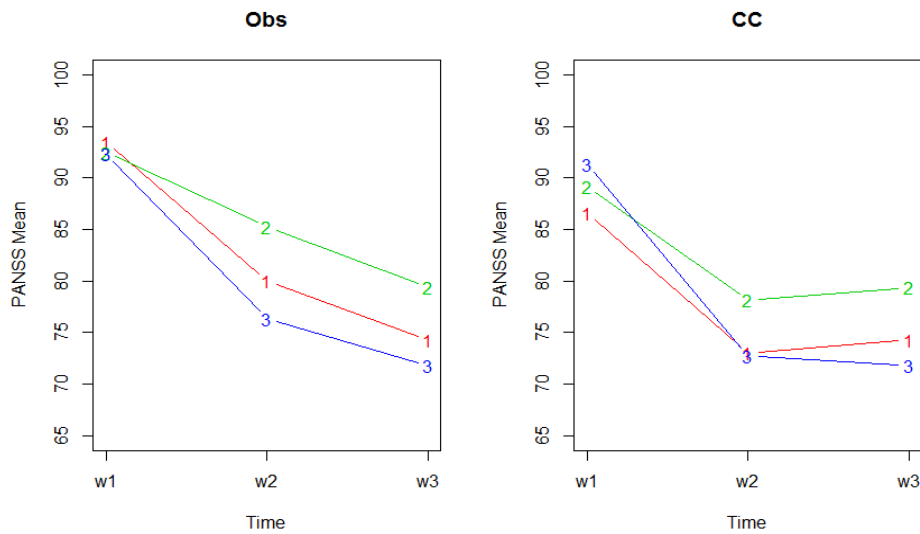
Table 2.19 shows the mean estimates of β for the different fitting methods using a saturated model with a different parameter for each of the treatment by time combinations. Table 2.20 shows the standard errors, which were calculated using the bootstrap based on 100 random samples. As seen before, the CC and IPW methods have higher standard errors than the other methods. Note that CC and Obs give the same results for β_7, β_8 and β_9 as explained earlier.

Table 2.19: Schizophrenia data: Estimates of coefficients β for different fitting methods.

	CC	Obs	IPW	LME	LI	LOCF
β_1	86.61	93.44	88.57	93.44	93.44	92.56
β_2	89.14	92.49	92.22	92.49	92.49	96.93
β_3	91.35	92.34	94.43	92.34	92.34	91.42
β_4	73.00	80.02	75.60	82.35	84.42	81.04
β_5	78.14	85.29	81.70	86.34	87.27	83.12
β_6	72.78	76.34	75.52	76.40	76.45	82.55
β_7	74.32	74.32	76.55	79.92	85.74	80.16
β_8	79.38	79.38	82.19	83.38	88.52	87.23
β_9	71.79	71.79	74.05	73.39	75.46	81.41

Table 2.20: Schizophrenia data: Standard error for estimates of coefficients β for different fitting methods.

	CC	Obs	IPW	LME	LI	LOCF
$se(\beta_1)$	2.78	1.98	2.99	1.68	1.91	1.98
$se(\beta_2)$	2.65	2.06	2.81	1.89	1.70	2.06
$se(\beta_3)$	1.45	0.98	1.44	1.00	1.16	0.98
$se(\beta_4)$	3.01	2.60	3.25	2.15	2.77	2.64
$se(\beta_5)$	2.74	2.49	2.92	1.96	2.75	2.29
$se(\beta_6)$	1.55	1.32	1.50	1.25	1.40	1.15
$se(\beta_7)$	2.95	2.95	3.17	2.48	3.32	2.51
$se(\beta_8)$	3.35	3.35	3.53	2.61	4.00	2.60
$se(\beta_9)$	1.68	1.68	1.67	1.42	1.52	1.09

Figures 2.5, 2.6 and 2.7 show estimated treatment effects and nominal 95% confidence intervals for the different fitting methods for times 1, 2 and 3 respectively. In all figures, the complete case value is lower than the others. This is consistent with the previous simulation results. There is high variability for the IPW again, but in Figure 2.7 it is even higher for the LI. In Figure 2.7 we note that Complete Case and Observed data give very similar estimates in all plots with IPW, LME, LI and LOCF progressively moving away.

Figure 2.5: Estimated parameters β_1, β_2 and β_3 and nominal 95% confidence intervals.
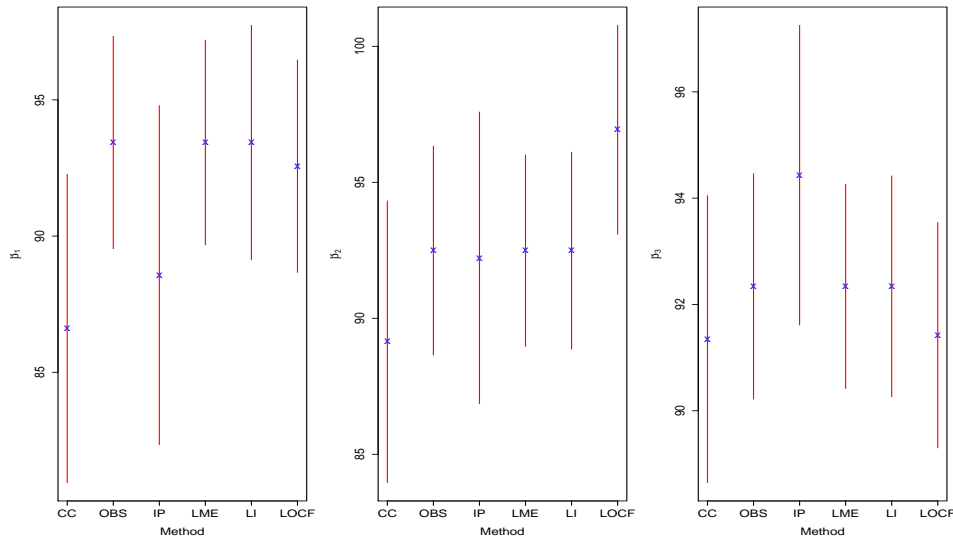
Figure 2.6: Estimated parameters β_4, β_5 and β_6 and nominal 95% confidence intervals.
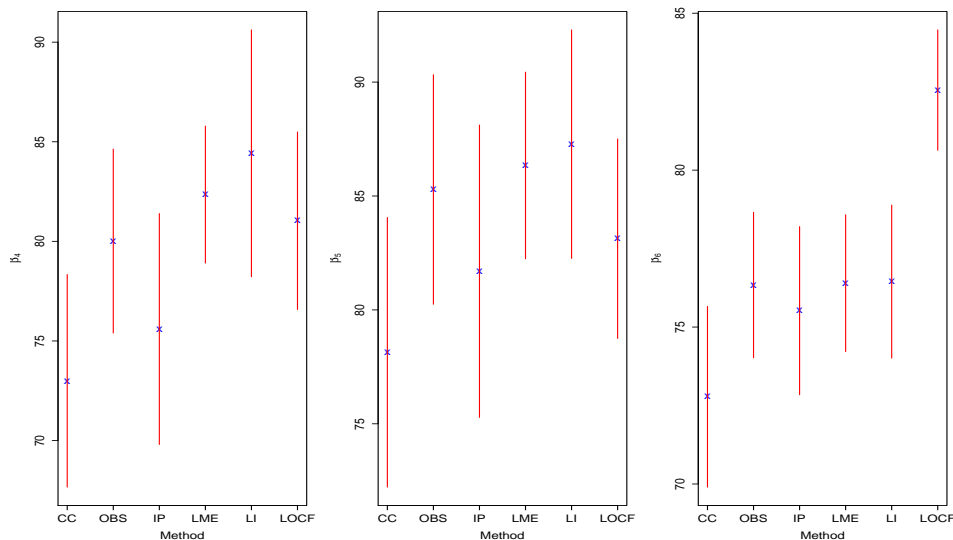
Figure 2.7: Estimated parameters β_7, β_8 and β_9 and nominal 95% confidence intervals.
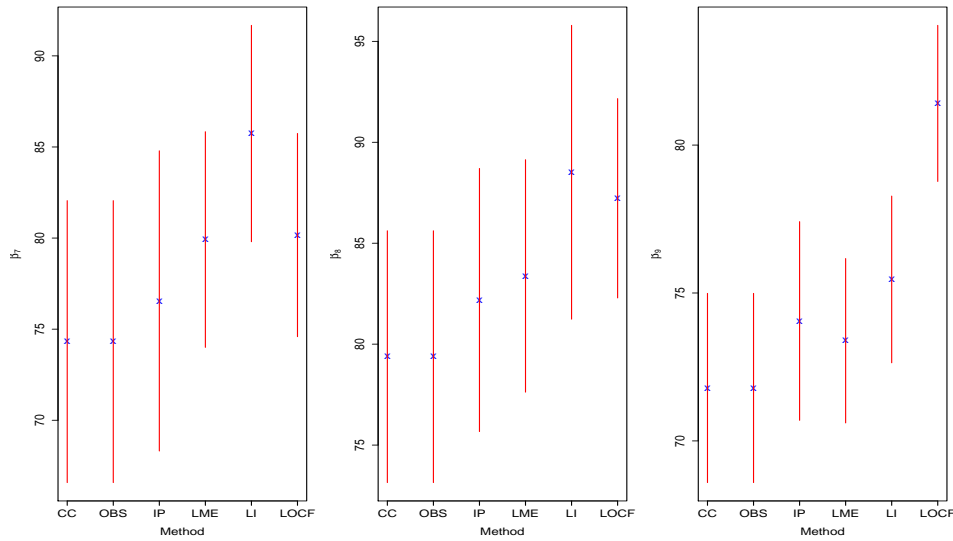
2.5 Conclusion and Discussion

In this chapter we described several different methods that can be used for longitudinal data under MCAR, MAR, MNAR and SP dropout assumptions. We found that only IPW and LME methods give consistent estimates under MAR, while LI is valid under SP. All of the methods fail to give consistent estimates under MNAR. In addition, we noted that the standard errors for the coefficients estimated by the IPW are significantly higher than for other methods.

We note that in CC and Obs methods there is repetition in the results for the second time in the two timepoints examples or the third time in case of three timepoints examples. That is because CC are people who make it to the final timepoint. So at that time the observed data are the complete case people. At earlier times the observed data are CC plus people who will dropout before the end. We will investigate the LI performance more in the next chapter.

Chapter 3

Performance of Linear Increment Method Under Shared Parameter, MAR and MNAR Dropout

3.1 Introduction

In this chapter we investigate the linear increment method (LI) and provide a discussion of the general model and its assumptions following Diggle et al. [2007]. Furthermore, we discuss the definition of independent censoring as introduced by Diggle et al. [2007]. We take the two timepoints situation as a special case and investigate how the LI method performs under shared parameter, MAR and MNAR dropout. Derivation and illustration of theoretical least false values are made under MAR and MNAR dropout. To check, we compare the theoretical least false values with simulation results for selected parameter combinations. An alternative technique was presented in Aalen and Gunnes [2010], who introduced a similar method that does not rely on a model for the missingness mechanism as we do. Using the linear increments model, they introduced a dynamic approach for reconstructing longitudinal data with missing observations.

In Section 3.2 we introduce the general model and the notation which we will use for all our analyses. Section 3.3 introduces the two timepoints situation as a special case. It includes brief details of how to use the extended skewed normal density to find the expectations needed. Section 3.4 presents the results from analyses of different simulation studies. Finally, in Section 3.5 we discuss all of our results and give some closing remarks.

3.2   The General Model

3.2.1   Assumptions

The linear increment model for analysing longitudinal data with dropouts was introduced by Farewell [2006]. Longitudinal data analysis methods often consider the marginal Generalized Estimating Equations (GEE) approach. However, this approach does not make efficient use of time ordering for the process. Diggle et al. [2007] and Aalen and Gunnes [2010] pointed out the advantages of using the linear increments models to analyse the dynamic structure of longitudinal data with dropout.

Using the notation found in Diggle et al. [2007], we denote the true history to time t by $\mathscr{G}_t = \{Y(s), X(s), \epsilon(s) : s = 1, \ldots, t\}$, which includes the response process Y, the covariates X and the measurement error process ϵ, up to and including time t. We will assume that the covariates $X(t)$ are obtained and are known immediately before time t, this timepoint could be $t - 1$ or at time 0 for those covariates known at baseline. Then the true history just before t is defined as:

$$\mathscr{G}_{t-} = \mathscr{G}_{t-1} \cup \{X(t)\} \tag{3.1}$$

which includes all information associated with Y, X and ϵ that are available strictly before time t. Thus we can say that $X(t)$ might include time varying covariates. We can also consider a dynamic process for the dropout. Let R_i be an indicator process for subject i, with $R_i(t) = 1$ if subject i has not dropped out at time t and is remaining in the study, and $R_i(t) = 0$ otherwise. We let $\mathscr{R}_t$ be the history of these indicator processes up to time t. And let $\mathscr{R}_{t-}$ to be the history of the dropout process immediately before time t.

We denote the history of the observed data X, Y and R as defined in Diggle et al. [2007] as

$$
\begin{aligned}
\mathscr{F}_t &= \{X(s), Y(s), R(s) : s = 1, ..., t\} \\
\mathscr{F}_{t-} &= \mathscr{F}_{t-1} \cup \{X(t), R(t)\}.
\end{aligned}
$$

The increments in $Y(t)$ are defined as $\Delta Y(t) = Y(t) - Y(t-1)$.

Diggle et al. [2007] proposed a linear increment model of the following form:

$$Y(t) = \sum_{s=1}^{t} X(s)\beta(s) + M(t) + \epsilon(t) \tag{3.2}$$

where $\epsilon(t)$ is a pure noise term, independent of other components, $M(t)$ is a random effect and $\beta(s)$ is a vector of time-varying effects. There are three key components to the model.

- Martingale error structure. Diggle et al. [2007] assume that the random effect is a martingale process. In words a martingale is a sequence of random variables (a

stochastic process) with the criterion that the conditional expectation of the next value in the sequence is equal to the present observed value, and this does not depend otherwise on earlier values. In general, $M(t)$ is a martingale, hence it meets the condition of $E[M(1)] = 0$ and $E[M(2)|M(1)] = M(1)$ and $E[M(t) - M(t-1)] = 0$.

- Linear model for changes in expected Y. We will consider this form for our models:

$$E[\Delta Y(t)|\mathscr{G}_{t-}] = E[Y(t) - Y(t-1)|\mathscr{G}_{t-}] \tag{3.3}$$

where at time t, the mean response of $Y(t)$ is

$$
\begin{aligned}
E[Y(t)|\mathscr{G}_{t-}] &= E[(\sum_{s=1}^{t} X(s)\beta(s) + M(t) + \epsilon(t))|\mathscr{G}_{t-}] \\
&= E[\sum_{s=1}^{t} X(s)\beta(s)|\mathscr{G}_{t-}] + E[M(t)|\mathscr{G}_{t-}] + E[\epsilon(t)|\mathscr{G}_{t-}] \\
&= X(1)\beta(1) + X(2)\beta(2) + \cdots \\
&+ X(t-1)\beta(t-1) + X(t)\beta(t) + E[M(t)|\mathscr{G}_{t-}] + E[\epsilon(t)|\mathscr{G}_{t-}].
\end{aligned}
$$

Note that $E[\epsilon(t)|\mathscr{G}_{t-}] = 0$ since there is no information available about $\epsilon(t)$ before time t. Similarly at time $t-1$

$$
\begin{aligned}
E[Y(t-1)|\mathscr{G}_{t-}] &= E[\sum_{s=1}^{t-1} X(s)\beta(s) + M(t-1) + \epsilon(t-1)] \\
&= E[\sum_{s=1}^{t-1} X(s)\beta(s)] + E[M(t-1)] + E[\epsilon(t-1)] \\
&= X(1)\beta(1) + X(2)\beta(2) + \cdots \\
&+ X(t-1)\beta(t-1) + M(t-1) + \epsilon(t-1).
\end{aligned}
$$

Thus, using the martingale structure that $E[M(t)|\mathscr{G}_{t-}] = M(t-1)$, we have

$$E[Y(t) - Y(t-1)|\mathscr{G}_{t-}] = X(t)\beta(t) - \epsilon(t-1) \tag{3.4}$$

where $\epsilon(t-1)$ is the measurement error at time $t-1$. Note that we have the measurement error on the left at $t-$ which is continuous time, while on the right we have $t-1$, as it is in discrete time. Simply we can say the mean difference of measurement error is $-\epsilon(t-1)$, since there is no information available about $\epsilon(t)$ before time t while earlier values are known through $\mathscr{G}_{t-}$. Diggle et al. [2007] put forward an idea which links the stochastic processes to the linear increment as the following: They state "Incremental models correspond, on the cumulative scale, to models where the residuals form a kind of random walk, which can be thought of as additional random effects. To see this, the notion of a transform from the theory of

discrete stochastic processes is required." For a reference in the *Theory of Discrete Stochastic Processes* see Williams [1991]. Define the cumulative regression functions, as in Diggle et al. [2007] as

$$B(t) = \sum_{s=1}^{t} \beta(s), \text{ with } B(0) = 0.$$

The transform of B by X, is given by:

$$X.B(t) = \sum_{s=1}^{t} X(s)\{B(s) - B(s-1)\} = \sum_{s=1}^{t} X(s)\beta(s). \tag{3.5}$$

It follows that the dynamic linear increments model (3.4) now can be written in the form

$$Y(t) = X.B(t) + M(t) + \epsilon(t) \tag{3.6}$$

where $X.B(t)$ is defined in equation (3.5).

- Independent censoring: The model meets the condition

$$E[\Delta Y(t)|\mathscr{G}_{t-}, \mathscr{R}_{t-}] = E[\Delta Y(t)|\mathscr{G}_{t-}]. \tag{3.7}$$

This means that additional knowledge of previous observation patterns does not affect the expected increment given that the time history of $X(t)$ and $Y(t)$ is known.

In the following we will work through the theory of Section 4 in Diggle et al. [2007] to build a model in our terminology for two timepoints as a special case. We will show that the model we use later in this chapter (Section 3.3) is a special case of the general model here. We will assume one time constant covariate x, and let Y be always observed at time 1, so that at the baseline we observe Y_1, R, and x. Moreover at time 2, if $R = 1$, then we observe Y_2 as well. Here

$$\begin{aligned}
\mathscr{G}_2 &= \{Y_1, Y_2, x, \epsilon_1, \epsilon_2\}, \mathscr{G}_1 = \{Y_1, x, \epsilon_1\}, \\
\mathscr{F}_2 &= \{Y_1, Y_2, x, R\}, \ \mathscr{F}_1 = \{Y_1, x\}, \ \mathscr{R}_2 = \{R\}.
\end{aligned} \tag{3.8}$$

For instance, at time 2; equation (3.7) will be : $E[\Delta Y(2)|\mathscr{G}_{2-}, \mathscr{R}_{2-}] = E[Y_2 - Y_1|x, Y_1, \epsilon_1, R]$. In our context x is a time constant scalar and the model in (3.2) for time 1 is:

$$Y_1 = \beta_1 + \beta_2 x + M_1 + \epsilon_1. \tag{3.9}$$

Then, the increment model defined in (3.4) is:

$$Y_2 - Y_1 = \gamma_1 + \gamma_2 x + (M_2 - M_1) + \epsilon_2 - \epsilon_1 \tag{3.10}$$

where $\gamma_1 = \beta_3 - \beta_1$ and $\gamma_2 = \beta_4 - \beta_2$.

Note M_1 can be found from Y_1, x, and ϵ_1, thus at time 2:

$$
\begin{aligned}
Y_2 &= (\beta_1 + \gamma_1) + (\beta_2 + \gamma_2)x + M_2 + \epsilon_2 \\
&= \beta_3 + \beta_4 x + M_2 + \epsilon_2.
\end{aligned}
\tag{3.11}
$$

In addition, at time 1: B transform x is $(x.B)(1) = \beta_1 + \beta_2 x$, hence the model defined in (3.6) will be: $Y_1 = \beta_1 + \beta_2 x + M_1 + \epsilon_1$, and this matches the result in equation (3.9). Similarly, at time 2: $(x.B)(2) = (\beta_1 + \gamma_1) + (\beta_2 + \gamma_2)x$, and the model defined in (3.6) will be: $Y_2 = (\beta_1 + \gamma_1) + (\beta_2 + \gamma_2)x + M_2 + \epsilon_2$. In other words, $Y_2 = \beta_3 + \beta_4 x + M_2 + \epsilon_2$, and this matches the result in equation (3.11). Note that we will discuss the special case when $M_2 = M_1$ in the next section, and we will assume $M_2 = M_1 = U \sim N(0, \sigma_U^2)$.

3.2.2 No missingness case

First we consider the case that there is no missingness, i.e. for now the complete case is considered. Following Diggle et al. [2007] Section 4.1.1 we assume we observe Y_2, then the true histories at time 0, 1, 2 respectively are as follows. At time 0, we have $\mathscr{G}_0 = \{x\}$. At time 1 we have $\mathscr{G}_1 = \{x, Y_1, \epsilon_1\}$. At time 2 $\mathscr{G}_2 = \{x, Y_1, Y_2, \epsilon_1, \epsilon_2\}$, where ϵ_1 and ϵ_2 are the measurement errors. Note that for simplicity, we set $Y(0) = \epsilon(0) = 0$. Since x is time-fixed then $X(t) = X(t-1) = x$, and the true history defined in (3.1) for this case (no missingness) is $\mathscr{G}_{t-} = \mathscr{G}_{t-1} \cup \{x\} = \mathscr{G}_{t-1}$. Hence the linear increment model as defined in (3.4) is: $E[\Delta Y(t)|\mathscr{G}_{t-}] = X(t)\beta(t) - \epsilon(t-1)$. Note that the unconditional expectation is $E[\Delta Y(t)] = X(t)\beta(t)$ as the expected error is zero.

For example: The general model for time 1:

$E[\Delta Y(1)|\mathscr{G}_0] = E[Y_1 - Y_0|\mathscr{G}_0] = E[Y_1|x] = X(1)\beta(1)$ where $\beta(1) = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$

and $X(1) = \begin{pmatrix} 1 & x \end{pmatrix}$, thus $E[\Delta Y(1)|\mathscr{G}_0] = \beta_1 + \beta_2 x$.

Similarly the general model for time 2 is:

$E[\Delta Y(2)|\mathscr{G}_1] = E[Y_{(2)} - Y_{(1)}|x, Y_1, \epsilon_1] = X(2)\beta(2) - \epsilon(1) = \gamma_1 + \gamma_2 x - \epsilon_1$

where $\mathscr{G}_1 = \{x, Y_1, \epsilon_1\}$, ϵ_1 is the measurement error, $\beta(2) = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} \beta_3 - \beta_1 \\ \beta_4 - \beta_2 \end{pmatrix}$

and $X(2) = \begin{pmatrix} 1 & x \end{pmatrix}$. This corresponds to Assumption 1. Again the unconditional expectation is $E[\Delta Y(2)] = \gamma_1 + \gamma_2 x$ as the expected error is zero.

3.2.3 The dropout case

Here we will illustrate the role of history $(\mathscr{G}, \mathscr{F}, \mathscr{R})$ defined previously in (3.8) now in the case of two timepoints. Suppose we always observe Y_1 in sequence $R_1 = 1$, and we might

observe Y_2 some times but not always because of dropout. Thus we must treat R_2 as a random variable, say $R_2 = R$, and the process is R_t. Hence $\mathscr{R}_1 = \{.\}$, $\mathscr{R}_2 = \{R\}$.

The true history will be the same as its definition in the previous section under the no missingness case because it does not contain R. The observed history at time 0, 1, 2 respectively is: At time 0 $\mathscr{F}_0 = \{x\}$, at time 1 $\mathscr{F}_1 = \{x, Y_1, R\}$, and at time 2 $\mathscr{F}_2 = \{x, Y_1, Y_2, R\}$. Note this is different from $\mathscr{G}$, as here there is no ϵ. In general, we will assume

$$E[\Delta Y(t)|\mathscr{G}_{t-}, \mathscr{R}_{t-}] = E[\Delta Y(t)|\mathscr{G}_{t-}].$$

It follows that

$$E[\Delta Y(t)|x, Y_1, \epsilon_1, R] = \gamma_1 + \gamma_2 x - \epsilon_1$$

and this corresponds to Assumption 3.

We want to find $E[\Delta Y(t)|\mathscr{F}_{t-}]$ from $E[\Delta Y(t)|\mathscr{G}_{t-}]$. In general

$$E_{\mathscr{G}_{t-}|\mathscr{F}_{t-}}[X(t)\beta(t) - \epsilon(t-1)|\mathscr{G}_{t-}] = E_{\epsilon|\mathscr{F}_{t-}}[X(t)\beta(t) - \epsilon(t-1)] = X(t)\beta(t) - E[\epsilon(t-1)|\mathscr{F}_{t-}].$$

For example at $t = 1$, $E[\Delta Y(1)|\mathscr{F}_{1-}] = E[Y_1|x] = \beta_1 + \beta_2 x - E[\epsilon_1|x]$, we have assumed ϵ_1 is independent of x, and $E[\epsilon_1|x] = 0$. Therefore, $E[Y_1|x] = \beta_1 + \beta_2 x$ as expected. Also at $t = 2$, $E[\Delta Y(2)|\mathscr{F}_{2-}] = \gamma_1 + \gamma_2 x - E[\epsilon_2|x, R, Y_1] = \gamma_1 + \gamma_2 x$, again as expected.

3.3 Two Timepoints As a Special Case

We will use the longitudinal model defined in the Introduction for two timepoints to generate the data and the missingness model. In the following we will explore the independent censoring assumption before continuing to investigate how the LI method performs under shared parameter, MAR and MNAR dropout.

3.3.1 Independent censoring

Here we investigate whether our models satisfy the independent censoring assumption (3.7). In the following will explore the SP, MAR and MNAR models and show if the independent censoring assumption is valid.

Shared parameter

The shared parameter dropout model is defined as:

$$P(R = 1|U) = \frac{e^{\theta_0 + \theta_1 U}}{1 + e^{\theta_0 + \theta_1 U}}$$

where U is a subject specific random effect. We want to investigate if this is independent censoring.

It is clear that in this model R depends on U, therefore knowing R tells us about U.

To be independent censoring, we have to verify (3.7) which says $E[\Delta Y(t)|\mathscr{G}_{t-}, \mathscr{R}_{t-}] = E[\Delta Y(t)|\mathscr{G}_{t-}]$. We attempt to rewrite this in our terminology. On the left hand side of (3.7) we have $\mathscr{G}_{2-} = \{x, Y_1, \epsilon_1\}$, $\mathscr{R}_{t-} = \{R\}$ then $E[\Delta Y(t)|\mathscr{G}_{t-}, \mathscr{R}_{t-}] = E[\Delta Y(t)|x, Y_1, \epsilon_1, R]$. Now, since R depends on U, therefore $E[Y_2 - Y_1|x, Y_1, \epsilon_1, R] = E[Y_2 - Y_1|x, Y_1, \epsilon_1, U, R]$, but here we assume U is known anyway. Then,

$$E[Y_2 - Y_1|x, Y_1, \epsilon_1, U, R] = E[Y_2 - Y_1|x, Y_1, \epsilon_1, U] = E[Y_2 - Y_1|x, Y_1, \epsilon_1]$$

and this matches the right hand side of (3.7). Thus ΔY is independent of R, and so it is independent censoring.

MAR

The MAR dropout model is defined as:

$$P(R = 1|Y_1) = \frac{e^{\theta_0 + \theta_1 Y_1}}{1 + e^{\theta_0 + \theta_1 Y_1}}.$$

We want to investigate if this is independent censoring.

We will adopt a similar approach to investigating the shared parameter. Here we know that R depends on Y_1. In other words, knowing Y_1 tells us about R, therefore

$$E[Y_2 - Y_1|x, Y_1, \epsilon_1, R] = E[Y_2 - Y_1|x, Y_1, \epsilon_1]$$

thus ΔY is conditionally independent of R, and so it is independent censoring.

MNAR

The MNAR dropout model is defined as:

$$P(R = 1|Y_1, Y_2) = \frac{e^{\theta_0 + \theta_1 Y_1 + \theta_2 Y_2}}{1 + e^{\theta_0 + \theta_1 Y_1 + \theta_2 Y_2}}.$$

We want to investigate if this is independent censoring.
Clearly R depends on Y_1 and Y_2, and hence this is not independent censoring because

$$E[Y_2 - Y_1|x, Y_1, \epsilon_1, R] \neq E[Y_2 - Y_1|x, Y_1, \epsilon_1].$$

3.3.2 Linear increment method with two timepoints

Instead of writing $\Delta Y(2)$, we will let $D_i = Y_{2i} - Y_{1i}$. We assume the correct model for responses, i.e.:

$$E[Y_{1i}] = \beta_1 + \beta_2 x_i \text{ and } E[Y_{2i}] = \beta_3 + \beta_4 x_i.$$

So for fully observed data

$$E[D_i] = (\beta_3 - \beta_1) + (\beta_4 - \beta_2)x_i = \gamma_1 + \gamma_2 x_i \quad \text{say.}$$

We will estimate (β_1, β_2) by ordinary least squares (OLS) of Y_1 on x, and (γ_1, γ_2) by OLS of D on x whenever $R = 1$. Then we take

$$\hat{\beta}_3 = \hat{\beta}_1 + \hat{\gamma}_1 \text{ and } \hat{\beta}_4 = \hat{\beta}_2 + \hat{\gamma}_2.$$

From simple linear regression we know

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n Y_{1i}(x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2} \text{ and } \hat{\beta}_1 = \overline{Y_1} - \hat{\beta}_2 \overline{x}.$$

The extra random variable R complicates the behavior at time 2, but we have

$$\hat{\gamma}_2 = \frac{\sum_{i=1}^n R_i D_i(x_i - \overline{Rx}/\overline{R})}{\sum_{i=1}^n R_i(x_i - \overline{Rx}/\overline{R})^2} \text{ and } \hat{\gamma}_1 = \frac{\overline{RD}}{\overline{R}} - \hat{\gamma}_2 \frac{\overline{Rx}}{\overline{R}}. \tag{3.12}$$

For example, the $\overline{Rx}/\overline{R}$ term arises as we use the mean x amongst observed values, which can be written as $\sum_{i=1}^n R_i x_i / \sum_{i=1}^n R_i$. Dividing top and bottom by n allows us to change the sums to means which is helpful for later use of the laws of large numbers (LLN), for reference see Ross [2009]. We will study what happens to the estimates under the three different dropout models, as n increases.

Now let $n \to \infty$. There are five fundamental variables: U, ϵ_1, ϵ_2, R and x. At times (including in this part) it may be more convenient to work with Y_1, Y_2 and/or D instead of U, ϵ_1 and ϵ_2. Note that $\hat{\beta}_1$ and $\hat{\beta}_2$ depend only on Y_{1i} and x_i, both of which are always observed, and we can deal with these without referring to the dropout model. We use a superscript $*$ to imply the limiting value. We will go through this in full, even though it is standard from linear regression, to illustrate the methods to come. We will use subscripts to indicate which variables we are considering when taking expectations, unless it is perfectly clear.

As the sample size increases, the sample means converge to the corresponding expected values. This implies that

$$
\begin{aligned}
\hat{\beta}_2 \to \beta_2^* &= \frac{E_{x,Y_1}\{Y_1(x - E_x[x])\}}{E_x\{(x - E_x[x])^2\}} = \frac{E_x\left[E_{Y_1|x}\{Y_1\}(x - E_x[x])\right]}{E_x\{(x - E_x[x])^2\}} \\
&= \frac{E_x\left[(\beta_1^G + \beta_2^G x)(x - E_x[x])\right]}{E_x\{(x - E_x[x])^2\}} \\
&= \beta_1^G \frac{E_x(x - E_x[x])}{E_x\{(x - E_x[x])^2\}} + \beta_2^G \frac{E_x\left[x(x - E_x[x])\right]}{E_x\{(x - E_x[x])^2\}} \\
&= \beta_2^G
\end{aligned}
$$

as $E_{Y_1|x}[Y_1] = \beta_1^G + \beta_2^G x$, $E_x(x - E_x[x]) = 0$ and $E_x\left[x(x - E_x[x])\right] = E_x\left\{(x - E_x[x])^2\right\}$.

Similarly

$$
\begin{aligned}
\hat{\beta}_1 \to \beta_1^* &= E_x\left\{E_{Y_1|x}[Y_1] - \beta_2^* E_x[x]\right\} = E_x\left\{\beta_1^G + \beta_2^G x - \beta_2^* E_x[x]\right\} \\
&= \beta_1^G + \beta_2^G E_x[x] - \beta_2^* E_x[x] = \beta_1^G.
\end{aligned}
$$

Hence for large samples our estimates at time 1 will converge to the correct (generating) values. To repeat, this is obvious but helps to illustrate the methods to come.

The coefficients in the model for D are of most interest. As n increases, the LLN implies that $\hat{\gamma}_2$ converges to γ_2^* given by

$$
\hat{\gamma}_2 \to \gamma_2^* = \frac{E\left\{RD(x - E[Rx]/E[R])\right\}}{E\left\{R(x - E[Rx]/E[R])^2\right\}} \tag{3.13}
$$

and

$$
\hat{\gamma}_1 \to \gamma_1^* = \frac{E[RD]}{E[R]} - \gamma_2^* \frac{E[Rx]}{E[R]}. \tag{3.14}
$$

To make the calculation easier, we can rewrite γ_2^* defined in (3.13) as:

$$
\gamma_2^* = \frac{E[RDx] - E[RD]E[Rx]/E[R]}{E[Rx^2] - (E[Rx])^2/E[R]}. \tag{3.15}
$$

3.3.3 Performance under shared parameter dropout

We start with equations (3.14) and (3.15). The shared parameter dropout model is

$$
P(R = 1|U) = \text{expit}\left(\theta_0 + \theta_1 U\right)
$$

and the expectation in equations (3.14) and (3.15) is with respect to the five variables: U, ϵ_1, ϵ_2, R and x.

First consider ϵ_1 and ϵ_2. Neither R nor x are affected by these and so if we take expectation with respect to them, only D is affected and

$$
E_{\epsilon_1, \epsilon_2 | x}\left\{D\right\} = \gamma_1^G + \gamma_2^G x.
$$

Note that we did not write $E_{\epsilon_1, \epsilon_2 | U, x}$ as ϵ_1, ϵ_2 are independent of U (and indeed x, but D depends on x so we left that in).

So

$$
\gamma_2^* = \gamma_1^G \frac{E\left\{R(x - E[Rx]/E[R])\right\}}{E\left\{R(x - E[Rx]/E[R])^2\right\}} + \gamma_2^G \frac{E\left\{Rx(x - E[Rx]/E[R])\right\}}{E\left\{R(x - E[Rx]/E[R])^2\right\}}
$$

where the expectation is now with respect to U, R and x. Under the SP model (1.9), R depends on U but not x. So $E[Rx] = E[R]E[x]$ and we can deal with the inner

expectations:

$$\gamma_2^* = \gamma_1^G \frac{E\left\{R(x - E[x])\right\}}{E\left\{R(x - E[x])^2\right\}} + \gamma_2^G \frac{E\left\{Rx(x - E[x])\right\}}{E\left\{R(x - E[x])^2\right\}}.$$

Once more R is independent of x and

$$
\begin{aligned}
\gamma_2^* &= \gamma_1^G \frac{E_{R,U}[R]E_x\left\{(x - E[x])\right\}}{E_{R,U}[R]E_x\left\{(x - E[x])^2\right\}} + \gamma_2^G \frac{E_{R,U}[R]E\left\{x(x - E[x])\right\}}{E_{R,U}[R]E_x\left\{(x - E[x])^2\right\}} \\
&= \gamma_2^G
\end{aligned}
$$

as required. Similarly

$$
\begin{aligned}
\hat{\gamma}_1 \to \gamma_1^* &= \frac{E[RD]}{E[R]} - \gamma_2^G \frac{E[Rx]}{E[R]} \\
&= E[D] - \gamma_2^G E[x] \\
&= \gamma_1^G + \gamma_2^G E[x] - \gamma_2^G E[x] \\
&= \gamma_1^G.
\end{aligned}
$$

So $\hat{\gamma}_1$ and $\hat{\gamma}_2$ converge to the correct values, and in turn so do $\hat{\beta}_3$ and $\hat{\beta}_4$ respectively, and this means that the linear increment method provides consistent estimates under the shared parameter dropout model.

3.3.4 Performance under MAR

The MAR dropout model is:

$$P(R = 1|Y_1) = \frac{e^{\theta_0 + \theta_1 Y_1}}{1 + e^{\theta_0 + \theta_1 Y_1}} = \text{expit}(\theta_0 + \theta_1 Y_1).$$

Recall the results in the previous chapter, where we found that the LI does not seem to work under the MAR assumption. Now we will investigate the bias that occurs in the parameters of the LI model. To see how the LI method works under the MAR model, we will apply the MAR assumptions to the equations (3.14) and (3.15). Under MAR, R depends on x through Y_1, so $E[Rx] \neq E[R]E[x]$. To make progress we can use an approximation based on the extended skew normal (ESN) to deal with the inner expectations. The definition of the ESN is found in the following section.

The Extended Skew Normal Distribution (ESN)

The ESN distribution is described in Johnson et al. [1995]. We will consider the definition and notation that was given in Ho et al. [2012]. A random variable w has an extended

skew normal distribution, $ESN(0, \sigma_w^2, \alpha, \nu)$, if it has density:

$$f(w) = \frac{\phi(w; 0, \sigma_w^2)\Phi(\alpha w + \nu)}{\Phi\left(\nu/\sqrt{1 + \alpha^2\sigma_w^2}\right)} \tag{3.16}$$

where $\phi(.; 0, \Sigma)$ is the normal density with mean 0 and dispersion Σ and $\Phi(.)$ denotes the distribution function of a standard normal. Note that $f(w)$ is a density function, then the integration of $f(w)$ over w equals 1 which implies:

$$\int_w \phi(w; 0, \sigma_w^2)\Phi(\alpha w + \nu)\, \mathrm{d}w = \Phi\left(\nu/\sqrt{1 + \alpha^2\sigma_w^2}\right). \tag{3.17}$$

The moment generating function (MGF) is:

$$M(t) = E[e^{t\,w}] = \frac{e^{\frac{1}{2}\sigma_w^2 t^2}\Phi\left[\frac{\alpha\sigma_w^2 t + \nu}{\sqrt{1 + \alpha^2\sigma_w^2}}\right]}{\Phi\left[\nu/\sqrt{1 + \alpha^2\sigma_w^2}\right]}. \tag{3.18}$$

Hence, the expectation of w is:

$$E[w] = M'(t)_{t=0} = \frac{\alpha\sigma_w^2}{\sqrt{1 + \alpha^2\sigma_w^2}}\frac{\phi(\bar{\nu})}{\Phi(\bar{\nu})}, \tag{3.19}$$

where $\bar{\nu} = \nu(1 + \alpha^2\sigma_w^2)^{-\frac{1}{2}}$.

The second moment is:

$$E[w^2] = M''(t)_{t=0} = \frac{\sigma_w^2\,\Phi(\bar{\nu}) - \bar{\nu}(\frac{\alpha\,\sigma_w^2}{\sqrt{1 + \alpha^2\sigma_w^2}})^2\phi(\bar{\nu})}{\Phi(\bar{\nu})}. \tag{3.20}$$

Further details are in Appendix A.1.

In the section below, we will be interested in integrals that can be written in terms of ESN. In particular

$$\int \phi(w; 0, \sigma_w^2)\Phi(\alpha w + \nu)\, \mathrm{d}w = \Phi\left(\nu/\sqrt{1 + \alpha^2\sigma_w^2}\right) = \Phi(\bar{\nu}). \tag{3.21}$$

$$\begin{aligned}
\int w\,\phi(w; 0, \sigma_w^2)\Phi(\alpha w + \nu)\, \mathrm{d}w &= \Phi\left(\nu/\sqrt{1 + \alpha^2\sigma_w^2}\right) E[w] = \Phi(\bar{\nu})E[w] \\
&= \Phi(\bar{\nu})\frac{\alpha\sigma_w^2}{\sqrt{1 + \alpha^2\sigma_w^2}}\frac{\phi(\bar{\nu})}{\Phi(\bar{\nu})} \\
&= \frac{\alpha\,\sigma_w^2\,\phi(\bar{\nu})}{\sqrt{1 + \alpha^2\sigma_w^2}}
\end{aligned} \tag{3.22}$$

and

$$
\begin{aligned}
\int w^2 \, \phi(w; 0, \sigma_w^2) \Phi(\alpha w + \nu) \, \mathrm{d}w \;\; &= \;\; \Phi\left(\nu/\sqrt{1 + \alpha^2 \sigma_w^2}\right) E[w^2] \\
&= \;\; \Phi(\bar{\nu}) E[w^2] \\
&= \;\; \Phi(\bar{\nu}) \frac{\sigma_w^2 \, \Phi(\bar{\nu}) - \bar{\nu}\left(\frac{\alpha \, \sigma_w^2}{\sqrt{1 + \alpha^2 \sigma_w^2}}\right)^2 \phi(\bar{\nu})}{\Phi(\bar{\nu})} \\
&= \;\; \sigma_w^2 \, \Phi(\bar{\nu}) - \bar{\nu}\left(\frac{\alpha \, \sigma_w^2}{\sqrt{1 + \alpha^2 \sigma_w^2}}\right)^2 \phi(\bar{\nu}). \qquad (3.23)
\end{aligned}
$$

3.3.5 Performance under MAR (continued) using ESN

Looking at the components of the equation (3.15), we need to find $E[R]$, $E[Rx]$, $E[RD]$, $E[Rx^2]$ and $E[RDx]$.

- As we know $E[R] = E[E[R|x]]$, thus we have to calculate $E[R|x]$ first, then we can find $E[R]$. Note that under MAR, R does not depend on ϵ_2 as it depends only on x, U and ϵ_1. We have

$$
\begin{aligned}
E[R|x] \;\; &= \;\; \int_{-\infty}^{\infty} P(R = 1 | x, \epsilon_1, U) f(\epsilon_1, U) \, \mathrm{d}\epsilon_1 \, \mathrm{d}U \\
&= \;\; \int_{-\infty}^{\infty} \mathrm{expit}\{\theta_0 + \theta_1 Y_1\} f(\epsilon_1, U) \, \mathrm{d}\epsilon_1 \, \mathrm{d}U \\
&= \;\; \int_{-\infty}^{\infty} \mathrm{expit}\{\theta_0 + \theta_1(\beta_1 + \beta_2 x) + \theta_1(U + \epsilon_1)\} f(\epsilon_1, U) \, \mathrm{d}\epsilon_1 \, \mathrm{d}U \\
&= \;\; \int_{-\infty}^{\infty} \mathrm{expit}\{K_1 + \theta_1 w_1\} f(w_1) \, \mathrm{d}w_1, \text{ say,}
\end{aligned}
$$

where

$$
\begin{aligned}
w_1 \;\; &= \;\; \epsilon_1 + U && (3.24) \\
K_1 \;\; &= \;\; \theta_0 + \theta_1(\beta_1 + \beta_2 x). && (3.25)
\end{aligned}
$$

We now use an approximation of the expit to the cumulative normal, see Johnson et al. [1995]:

$$
\mathrm{expit}(z) \approx \Phi(c\, z), \; c = \frac{16\sqrt{3}}{15\pi}. \qquad (3.26)
$$

Therefore:

$$
E[R|x] \;\; \approx \;\; \int_{-\infty}^{\infty} \Phi\{c(K_1 + \theta_1 w_1)\} f(w_1) \, \mathrm{d}w_1.
$$

From equation (3.24) $w_1 = \epsilon_1 + U$ and we assumed in our model in (1.1) that $\epsilon_1 \sim N(0, \sigma_{\epsilon_1}^2)$ and $U \sim N(0, \sigma_U^2)$. Hence we can say that w_1 is normally distributed with mean 0 and variance $\sigma_{w_1}^2$, i.e $w_1 \sim N(0, \sigma_{w_1}^2)$, where $\sigma_{w_1}^2 = \sigma_{\epsilon_1}^2 + \sigma_U^2$, which

allows us to replace $f(w_1)$ with $\phi(w_1; 0, \sigma_{w_1}^2)$ and

$$E[R|x] \quad \approx \quad \int_{-\infty}^{\infty} \Phi\{c(K_1 + \theta_1 w_1)\}\phi(w_1; 0, \sigma_{w_1}^2)\,\mathrm{d}w_1.$$

The integral is now equivalent to the numerator in (3.16) i.e. the pdf of ESN. We can thus use equation (3.21) to obtain

$$E[R|x] \quad \approx \quad \Phi\left[\frac{cK_1}{\sqrt{1 + c^2\theta_1^2\sigma_{w_1}^2}}\right]$$

$$= \quad \Phi\left[\frac{c(\theta_0 + \theta_1(\beta_1 + \beta_2 x))}{\sqrt{1 + c^2\theta_1^2\sigma_{w_1}^2}}\right]$$

$$= \quad \Phi(A_1 + A_2 x) \qquad (3.27)$$

where we have used equation (3.25) to replace K_1 and we have defined $A_1 = \frac{c}{\sqrt{1 + c^2\theta_1^2\sigma_{w_1}^2}}(\theta_0 + \theta_1\beta_1)$ and $A_2 = \frac{c}{\sqrt{1 + c^2\theta_1^2\sigma_{w_1}^2}}(\theta_1\beta_2)$.

Now we can find $E[R]$ by integrating the above expectation in (3.27) over x. Recall that $x \sim N(0, \sigma_x^2)$ so

$$E[R] \quad = \quad E_x[E[R|x]]$$

$$\approx \quad \int_{-\infty}^{\infty} \Phi(A_1 + A_2 x)\phi(x; 0, \sigma_x^2)\mathrm{d}x$$

then we use equation (3.21) to obtain

$$E[R] \quad \approx \quad \Phi\left[\frac{A_1}{\sqrt{1 + (A_2\sigma_x)^2}}\right]$$

$$= \quad \Phi(\nu_1) \qquad (3.28)$$

where $\nu_1 = \frac{A_1}{\sqrt{1 + (A_2\sigma_x)^2}}$.

- We can calculate $E[Rx]$, using

$$E[Rx] \quad = \quad E_x[xE[R|x]]$$

$$\approx \quad \int_{-\infty}^{\infty} x\,\Phi\left\{A_1 + A_2 x\right\}\phi(x; 0, \sigma_x^2)\mathrm{d}x.$$

We can use formula (3.22) to perform the integration, and the result obtained is:

$$E[Rx] \quad = \quad \frac{\sigma_x^2 A_2}{\sqrt{1 + A_2^2\sigma_x^2}}\phi\left(\frac{A_1}{\sqrt{1 + A_2^2\sigma_x^2}}\right) = \frac{\sigma_x^2 A_2}{\sqrt{1 + A_2^2\sigma_x^2}}\phi(\nu_1). \qquad (3.29)$$

- We now turn to $E[RD]$, where $D = Y_2 - Y_1 = \gamma_1 + \gamma_2 x + \epsilon_2 - \epsilon_1$. Clearly

$$
\begin{aligned}
E[RD] &= E[R(\gamma_1 + \gamma_2 x + \epsilon_2 - \epsilon_1)] \\
&= \gamma_1 E[R] + \gamma_2 E[Rx] + E[R\epsilon_2] - E[R\epsilon_1].
\end{aligned}
\tag{3.30}
$$

Note that under MAR, we have $E[R\epsilon_2] = 0$, so

$$
E[RD] = \gamma_1 E[R] + \gamma_2 E[Rx] - E[R\epsilon_1].
\tag{3.31}
$$

We need to calculate $E[R\epsilon_1] = E_{\epsilon_1}[\epsilon_1 E[R|\epsilon_1]]$. Consider first

$$
\begin{aligned}
E[R|\epsilon_1] &= \int_{-\infty}^{\infty} P(R = 1|x, \epsilon_1, U) f(x, U) \, dx \, dU \\
&= \int_{-\infty}^{\infty} \operatorname{expit}\{\theta_0 + \theta_1 Y_1\} f(x, U) \, dx \, dU \\
&= \int_{-\infty}^{\infty} \operatorname{expit}\{\theta_0 + \theta_1(\beta_1 + \beta_2 x + U + \epsilon_1)\} f(x, U) \, dx \, dU.
\end{aligned}
$$

Let $w_2 = K_2 + \theta_1 \epsilon_1$ and $w_3 = \beta_2 x + U$, where $K_2 = \theta_0 + \theta_1 \beta_1$. Then the variance of w_3 is $\sigma_{w_3}^2 = \beta_2^2 \sigma_x^2 + \sigma_U^2$, and the above equation can be rewritten as:

$$
E[R|\epsilon_1] = \int_{-\infty}^{\infty} \operatorname{expit}\{w_2 + \theta_1 w_3\} f(w_3) \, dw_3.
$$

We now approximate the $\operatorname{expit}(z)$ by $\{\Phi(cz)\}$, and since $w_3 \sim N(0, \sigma_{w_3}^2)$ we can say:

$$
\begin{aligned}
E[R|\epsilon_1] &\approx \int_{-\infty}^{\infty} \Phi\{c(w_2 + \theta_1 \, w_3)\} \phi(w_3; 0, \sigma_{w_3}^2) \, dw_3 \\
&= \Phi\left[\frac{c \, w_2}{\sqrt{1 + c^2 \theta_1^2 \sigma_{w_3}^2}} \right].
\end{aligned}
$$

Now re arranging the quantity inside we write

$$
E[R|\epsilon_1] \approx \Phi\left[A_3 + A_4 \epsilon_1\right]
$$

where $A_3 = \frac{c(\theta_0 + \theta_1 \beta_1)}{\sqrt{1 + c^2 \theta_1^2 \sigma_{w_3}^2}}$ and $A_4 = \frac{c \theta_1}{\sqrt{1 + c^2 \theta_1^2 \sigma_{w_3}^2}}$.

Now we want to find $E[R\epsilon_1]$. This is

$$
\begin{aligned}
E[R\epsilon_1] &= E_{\epsilon_1}[\epsilon_1 E[R|\epsilon_1]] \\
&\approx E_{\epsilon_1}[\epsilon_1 \Phi[A_3 + A_4\epsilon_1]] \\
&= \int \epsilon_1 \Phi[A_3 + A_4\epsilon_1] f(\epsilon_1) \, \mathrm{d}\epsilon_1 \\
&= \int \epsilon_1 \Phi[A_3 + A_4\epsilon_1] \phi(\epsilon_1; 0, \sigma_{\epsilon_1}^2) \, \mathrm{d}\epsilon_1 \\
&= \frac{\sigma_{\epsilon_1}^2 A_4}{\sqrt{1 + A_4 \sigma_{\epsilon_1}^2}} \phi\left(\frac{A_3}{\sqrt{1 + A_4^2 \sigma_{\epsilon_1}^2}}\right) \\
&= \frac{\sigma_{\epsilon_1}^2 A_4}{\sqrt{1 + A_4 \sigma_{\epsilon_1}^2}} \phi(\nu_2)
\end{aligned}
\tag{3.32}
$$

where we have used (3.22) and we have defined $\nu_2 = \frac{A_3}{\sqrt{1 + A_4^2 \sigma_{\epsilon_1}^2}}$. From equations (3.28), (3.29), and (3.32), we can now calculate the components required to build up equation (3.31) to get the value of $E[RD]$ as will be shown later.

- We turn now to $E[RDx]$. We have

$$
\begin{aligned}
E[RDx] &= E[Rx(\gamma_1 + \gamma_2 x + \epsilon_2 - \epsilon_1)] \\
&= \gamma_1 E[Rx] + \gamma_2 E[Rx^2] + E[Rx\epsilon_2] - E[Rx\epsilon_1]
\end{aligned}
$$

and again, under MAR $E[Rx\epsilon_2] = 0$. Then

$$
E[RDx] = \gamma_1 E[Rx] + \gamma_2 E[Rx^2] - E[Rx\epsilon_1].
\tag{3.33}
$$

It follows that we still need also to calculate $E[Rx^2]$ and $E[Rx\epsilon_1]$. We start with:

$$
\begin{aligned}
E[Rx^2] &= E_x[x^2 E[R|x]] \\
&\approx \int_{-\infty}^{\infty} x^2 \Phi\{A_1 + A_2 x\} \phi(x; 0, \sigma_x^2) \mathrm{d}x.
\end{aligned}
$$

Using form (3.23)

$$
E[Rx^2] \approx \Phi[\nu_1] - \nu_1 \left(\frac{A_2 \sigma_x^2}{\sqrt{1 + A_2^2 \sigma_x^2}}\right)^2 \phi[\nu_1].
\tag{3.34}
$$

- To calculate $E[Rx\epsilon_1]$ we need several stages. First

$$
E[Rx\epsilon_1] = \int \int \int x \, \epsilon_1 \exp\mathrm{it}\{\theta_0 + \theta_1(\beta_1 + \beta_2 x + U + \epsilon_1)\} f(U) f(\epsilon_1) f(x) \mathrm{d}U \, \mathrm{d}\epsilon_1 \, \mathrm{d}x.
$$

Using the approximation of the expit, then

$$E[Rx\epsilon_1] \quad \approx \quad \int \int \int x\,\epsilon_1\, \Phi\{c(\theta_0 + \theta_1(\beta_1 + \beta_2 x + U + \epsilon_1))\}\, f(U)\, f(\epsilon_1)\, f(x)\mathrm{d}U\, \mathrm{d}\epsilon_1\, \mathrm{d}x.$$

First we will integrate out U,

$$E[Rx\epsilon_1] \approx \int \int x\,\epsilon_1 \left\{ \int \Phi\{c(\theta_0 + \theta_1(\beta_1 + \beta_2 x + U + \epsilon_1))\}\, f(U)\, \mathrm{d}U \right\} f(\epsilon_1)\, f(x)\, \mathrm{d}\epsilon_1\, \mathrm{d}x.$$

(3.35)

We consider the inner integral, with ϵ_1 and x treated as constants:

$$\int \Phi\{c(\theta_0 + \theta_1(\beta_1 + \beta_2 x + U + \epsilon_1))\}\, f(U)\, \mathrm{d}U \quad = \quad \int \Phi\{K_3 + K_4 U\}\, f(U)\, \mathrm{d}U$$

where $K_3 = c(\theta_0 + \theta_1(\beta_1 + \beta_2 x + \epsilon_1))$ and $K_4 = c\theta_1$. Since $U \sim N(0, \sigma_U^2)$ we can write this as

$$\int \Phi\{c(\theta_0 + \theta_1(\beta_1 + \beta_2 x + U + \epsilon_1))\}\, f(U)\, \mathrm{d}U \quad = \quad \int \Phi\{K_3 + K_4 U\}\, \phi(U; 0, \sigma_U^2)\, \mathrm{d}U.$$

This is of the form (3.21) and so

$$\int \Phi\{c(\theta_0 + \theta_1(\beta_1 + \beta_2 x + U + \epsilon_1))\}\, f(U)\, \mathrm{d}U \quad = \quad \Phi\left[\frac{K_3}{\sqrt{1 + (K_4 \sigma_U^2)^2}}\right].(3.36)$$

We will now consider ϵ_1. Equation (3.36) allows (3.35) to be written as:

$$E[Rx\epsilon_1] \quad \approx \quad \int \int x\,\epsilon_1\, \Phi\left[\frac{K_3}{\sqrt{1 + (K_4 \sigma_U^2)^2}}\right] f(\epsilon_1) f(x)\, \mathrm{d}\epsilon_1\, \mathrm{d}x.$$

Replacing K_3 by its value and letting $K_5 = \frac{c(\theta_0 + \theta_1\beta_1) + c\theta_1\beta_2 x}{\sqrt{1 + (K_4\sigma_U)^2}}$ and $K_6 = \frac{c\theta_1}{\sqrt{1 + (K_4\sigma_U)^2}}$, then

$$\begin{aligned} E[Rx\epsilon_1] \quad &\approx \quad \int \int x\,\epsilon_1\, \Phi\{K_5 + K_6\epsilon_1\} f(\epsilon_1) f(x)\, \mathrm{d}\epsilon_1\, \mathrm{d}x \\ &= \quad \int x \left\{ \int \epsilon_1\, \Phi\{K_5 + K_6\epsilon_1\}\, f(\epsilon_1)\, \mathrm{d}\epsilon_1 \right\} f(x)\, \mathrm{d}x. \end{aligned}$$ (3.37)

Again we consider the inner integral, with x now treated as constant. Since $\epsilon_1 \sim N(0, \sigma_{\epsilon_1}^2)$ we can write this as

$$\int \epsilon_1\, \Phi\{K_5 + K_6\epsilon_1\}\, f(\epsilon_1)\, \mathrm{d}\epsilon_1 \quad = \quad \int \epsilon_1\, \Phi\{K_5 + K_6\epsilon_1\}\, \phi(\epsilon_1; 0, \sigma_{\epsilon_1}^2)\, \mathrm{d}\epsilon_1.$$

This matches the form (3.22), thus

$$
\begin{aligned}
\int \epsilon_1 \, \Phi\{K_5 + K_6\epsilon_1\} \, f(\epsilon_1) \, \mathrm{d}\epsilon_1 &= \frac{K_6 \, \sigma_{\epsilon_1}^2}{\sqrt{1 + (K_6 \, \sigma_{\epsilon_1})^2}} \, \phi\left(\frac{K_5}{\sqrt{1 + (K_6 \, \sigma_{\epsilon_1})^2}}\right) \\
&= K_7 \, \phi(K_8 + K_9 x) \quad\quad (3.38)
\end{aligned}
$$

where $K_7 = \frac{K_6 \, \sigma_{\epsilon_1}^2}{\sqrt{1+(K_6 \, \sigma_{\epsilon_1})^2}}$, $K_8 = \frac{c(\theta_0+\theta_1\beta_1)}{\sqrt{1+(K_4\sigma_U)^2}}$ and $K_9 = \frac{c\theta_1\beta_2 x}{\sqrt{1+(K_4\sigma_U)^2}}$. Now we will go back to (3.37) using the result in (3.38):

$$
E[Rx\epsilon_1] \approx K_7 \int x \, \phi(K_8 + K_9 x) f(x) \mathrm{d}x.
$$

Since $x \sim N(0, \sigma_x^2)$, thus

$$
E[Rx\epsilon_1] = K_7 \int x \, \phi(K_8 + K_9 x) \phi(x; 0, \sigma_x^2) \mathrm{d}x.
$$

Now we will use the result of Appendix A.2 to find the above integration. It follows that:

$$
E[Rx\epsilon_1] = -K_7 \frac{K_8 K_9}{(K_9^2 + 1)^{\frac{3}{2}}} e^{-\frac{1}{2}K_{10}} = A_5 \quad\quad (3.39)
$$

where $K_{10} = \frac{K_8^2}{K_9^2+1}$.

We now have all the terms needed for the limiting value γ_2^* defined in equation (3.15) and γ_1^* defined in equation (3.14), which can be calculated from equations (3.33), (3.31), (3.29), (3.28) and (3.34). We do not write the full expression because of the length. But to summarise:

$$
\begin{aligned}
E[RDx] &= \gamma_1 E[Rx] + \gamma_2 E[Rx^2] - E[Rx\epsilon_1] \\
&\approx \gamma_1 \frac{\sigma_x^2 A_2}{\sqrt{1 + A_2^2\sigma_x^2}} \phi(\nu_1) + \gamma_2 \left\{ \Phi[\nu_1] - \nu_1 \left(\frac{A_2 \, \sigma_x^2}{\sqrt{1 + A_2^2\sigma_x^2}}\right)^2 \phi[\nu_1] \right\} - A_5. \quad (3.40) \\
E[RD] &= \gamma_1 E[R] + \gamma_2 E[Rx] - E[R\epsilon_1] \\
&\approx \gamma_1 \Phi(\nu_1) + \gamma_2 \left\{ \frac{\sigma_x^2 A_2}{\sqrt{1 + A_2^2\sigma_x^2}} \phi(\nu_1) \right\} - \left(\frac{\sigma_{\epsilon_1}^2 A_4}{\sqrt{1 + A_4\sigma_{\epsilon_1}^2}} \phi(\nu_2)\right). \quad (3.41) \\
E[Rx] &= \frac{\sigma_x^2 A_2}{\sqrt{1 + A_2^2\sigma_x^2}} \phi(\nu_1). \quad\quad (3.42) \\
E[R] &\approx \Phi(\nu_1). \quad\quad (3.43) \\
E[Rx^2] &\approx \Phi[\nu_1] - \nu_1 \left(\frac{A_2 \, \sigma_x^2}{\sqrt{1 + A_2^2\sigma_x^2}}\right)^2 \phi[\nu_1]. \quad\quad (3.44)
\end{aligned}
$$

Then γ_2^* can be calculated from (3.40) to (3.44). We will investigate numerically in Section 3.4. Before then we look at performance under MNAR.

3.3.6   Performance under MNAR

To see how the LI method works under the MNAR model defined in (1.7), we will apply the MNAR assumptions to the equations (3.15) and (3.14), where the expectation is with respect to the four variables: ϵ_1, ϵ_2, R and x.

Under the MNAR dropout model, R depends on x. So $E[Rx] \neq E[R]E[x]$ and we have to use the approximation to the extended skew normal (ESN) to deal with the inner expectations.

For simplicity, we choose $\theta_1 = 0$, so we do not need to worry about Y_1, which will make the integrals easier for this description. Calculations for general θ_1 have been completed (and are used later in the thesis).

We need to calculate $E[R]$, $E[R\,x]$, $E[R\,x^2]$, $E[RD]$, $E[RD\,x]$. In turn,

$$E[RD] = E[R(\gamma_1 + \gamma_2 x + \epsilon_2 - \epsilon_1)] = \gamma_1 E[R] + \gamma_2 E[Rx] + E[R\epsilon_2] - E[R\epsilon_1] \qquad (3.45)$$

and

$$E[RDx] = E[Rx(\gamma_1 + \gamma_2 x + \epsilon_2 - \epsilon_1)] = \gamma_1 E[Rx] + \gamma_2 E[Rx^2] + E[Rx\epsilon_2] - E[Rx\epsilon_1] \quad (3.46)$$

meaning we need $E[R]$, $E[Rx]$, $E[Rx^2]$, $E[R\epsilon_2]$, $[Rx\epsilon_2]$, $E[R\epsilon_1]$, $[Rx\epsilon_1]$. Because we have taken $\theta_1 = 0$, R is independent of ϵ_1 and so $E[R\epsilon_1] = [Rx\epsilon_1] = 0$, and these terms are easily dealt with. In order to obtain the other terms, we will consider nested expectations beginning with R conditioned upon x and ϵ_2, just as in the previous section. Then we can use similar methods to those for the MAR assumption. Details are omitted but the final

results are:

$$
\begin{aligned}
E[R|x, \epsilon_2] &\approx \Phi\left[\frac{cF_1}{\sqrt{1+(c\theta_2\sigma_U)^2}}\right] \\
E[R|\epsilon_2] &\approx \Phi\left[F_4 + F_5\epsilon_2\right] \\
E[R] &\approx \Phi\left[\frac{F_4}{\sqrt{1+(F_5\sigma_{\epsilon_2})^2}}\right] \\
E[R|x] &\approx \Phi\left[\frac{F_6}{\sqrt{1+(F_7\sigma_{\epsilon_2})^2}}\right]
\end{aligned}
\tag{3.47}
$$

$$
E[Rx] = \frac{\sigma_x^2 F_9}{\sqrt{1+(F_9\sigma_x)^2}}\phi(\bar{\nu}_2)
\tag{3.48}
$$

$$
E[Rx^2] \approx -\bar{\nu}_2\left(\frac{F_9\sigma_x^2}{\sqrt{1+(F_9\sigma_x)^2}}\right)^2\phi(\bar{\nu}_2) + \sigma_x^2\Phi(\bar{\nu}_2)
\tag{3.49}
$$

$$
E[R\epsilon_2] = \frac{\sigma_x^2 F_5}{\sqrt{1+(F_5\sigma_{\epsilon_2})^2}}\phi(\bar{\nu}_1)
\tag{3.50}
$$

$$
E[R\,\epsilon_2\,x] = -K\frac{B_5\,B_6}{(B_6+1)^{3/2}}e^{-\frac{1}{2}\frac{B_5}{B_6^2+1}}
\tag{3.51}
$$

where, $F_1 = \theta_0 + \theta_2(\beta_3 + \beta_4 x + \epsilon_2)$, $F_3 = c\theta_2\beta_4$, $F_4 = \frac{c(\theta_0 + \theta_2\beta_3)}{\sqrt{1+(F_3\sigma_x)^2}}$, $F_5 = \frac{c\theta_2}{\sqrt{1+(F_3\sigma_x)^2}}$, $\bar{\nu}_1 = \frac{F_4}{\sqrt{1+(F_5\sigma_{\epsilon_2})^2}}$, $\bar{\nu}_2 = \frac{F_8}{\sqrt{1+(F_9\sigma_x)^2}}$, $F_8 = \frac{c}{\sqrt{1+(F_7\sigma_{\epsilon_2})^2}\sqrt{1+(c\theta_2\sigma_U)^2}}[\theta_0 + \theta_2\beta_3]$ and $F_9 = \frac{c}{\sqrt{1+(F_7\sigma_{\epsilon_2})^2}\sqrt{1+(c\theta_2\sigma_U)^2}}[\theta_2\beta_4]$.

Therefore, equation (3.45) will be:

$$
E[RD] \approx \gamma_1\Phi\left[\frac{F_4}{\sqrt{1+(F_5\sigma_{\epsilon_2})^2}}\right] + \gamma_2\left(\frac{\sigma_x^2 F_9}{\sqrt{1+(F_9\sigma_x)^2}}\phi(\bar{\nu}_2)\right) + \left(\frac{\sigma_x^2 F_5}{\sqrt{1+(F_5\sigma_{\epsilon_2})^2}}\phi(\bar{\nu}_1)\right).
\tag{3.52}
$$

Similarly, equation (3.46) will be:

$$
\begin{aligned}
E[RDx] \approx &\gamma_1\left(\frac{\sigma_x^2 F_9}{\sqrt{1+(F_9\sigma_x)^2}}\phi(\bar{\nu}_2)\right) + \gamma_2\left(-\bar{\nu}_2\left(\frac{F_9\sigma_x^2}{\sqrt{1+(F_9\sigma_x)^2}}\right)^2\phi(\bar{\nu}_2) + \sigma_x^2\Phi(\bar{\nu}_2)\right) \\
&+ \left(-K\frac{B_5\,B_6}{(B_6+1)^{3/2}}e^{-\frac{1}{2}\frac{B_5}{B_6^2+1}}\right).
\end{aligned}
\tag{3.53}
$$

Now, we have all of the components of (3.45)and (3.46) needed to find the limiting values γ_1^* and γ_2^* defined in equations (3.15) and (3.14).

In the next section we will show the numerical results of the work which has been done in theory to investigate the performance of the LI under MAR and MNAR.

3.4   Numerical Investigation

We will undertake a numerical investigation in two parts. In the first part we want to check that the theory for our least false values is correct. In the second part we explore in detail how the misspecification affects the γ estimates, using the theoretical results. Now that we have working formulas there is no need for simulations here. For example, we produce contour plots to show how the γs change when the MAR or MNAR coefficients θ are changed. First we show the limiting values γ_1^* and γ_2^* calculated from equations (3.40) to (3.44) for MAR work, and from equations (3.47) to (3.53) in the MNAR section, and demonstrate through simulations that they are correct by comparing the expected value which is calculated from the formulas mentioned above with the observed values which are calculated from equation (3.12). We generated a scalar $N(0,1)$ variable x, then we generated the longitudinal means $\mu_1 = \beta_1 + \beta_2 x$, $\mu_2 = \beta_3 + \beta_4 x$. This was followed by (Y_1, Y_2) from a bivariate normal distribution with mean (μ_1, μ_2). Missingness was generated from equations (1.5), (1.7) and (1.10) for the models MAR, MNAR and SP respectively. In all of the following simulations, unless it is stated otherwise, we take the parameters $\sigma_x = \sigma_{\epsilon_1} = \sigma_{\epsilon_2} = \sigma_U = \sqrt{0.5}$, $\beta = (-2, -2, -1, -1)$ (and then $\gamma_1 = \gamma_2 = 1$), $(\theta_0^{sp}, \theta_1^{sp}) = (0, \sqrt{0.5})$ and for MNAR we choose $\theta_1^{MN} = 0$, so we do not worry about Y_1, which has made the integrals in the theoretical parts easier. We will choose the MAR and MNAR parameters θ to give a similar dropout rate to SP. This choice of parameters will keep the dropout percentage always at about 50%. See Appendix A.3.

First we want to know at what n can we be confident that the theory is reasonable. We compare the theoretical results with the simulated as n increases, with results shown in Table 3.1 and Table 3.2. Based on this result for each set of parameters we studied sample size of 50000. In addition, 95% nominal confidence intervals (CI) for the expectations under SP, MAR and MNAR are shown in Table 3.3, Table 3.4 and Table 3.5 respectively.

Secondly, we will try to find several combinations of $(\theta_0, \theta_1, \theta_2)$ under fixed β and for given variances $\sigma_x^2 = 1$, $\sigma_U = \sigma_{\epsilon_1} = \sigma_{\epsilon_2} = \sqrt{0.5}$ to keep the total amount of the dropout fixed at about 50%. We will study the effect of changing θs. Table 3.10 and Table 3.11 show this effect. In each table the absolute value of the bias in the limiting values γ_1^* and γ_2^* is calculated. Then, for a given θ (and dropout), we will study the effect of changing the variance parameters. The result are shown in Table 3.12 and Table 3.13.

Finally we produced contour plots shown in Figure 3.1 to Figure 3.4

3.4.1   Comparing the theoretical results with the simulated as n increases

Here we attempt to show what happens as sample size increases, i.e. at what n can we be confident that the theory is reasonable. Results at $n = 50000$ show that the simulated

means all match the theoretical approximations very well, which supports the use of the theory. The simulations are denoted by for example $\bar{R}$, and the least false approximations by for example $E[R]$.

Table 3.1: Comparison of simulated and least false values for $n = 1000$

n		SP	MAR	MNAR
1000	$\bar{R}$	0.49	0.49	0.51
	$E[R]$	0.50	0.50	0.50
	$\overline{Rx}$	-0.01	-0.20	-0.09
	$E[Rx]$	0.00	-0.20	-0.11
	$\overline{RD}$	0.46	0.22	0.46
	$E[RD]$	0.50	0.25	0.45
	$\overline{RDx}$	0.40	0.31	0.41
	$E[RDx]$	0.50	0.30	0.39
	$\overline{Rx^2}$	0.45	0.51	0.48
	$E[Rx^2]$	0.50	0.50	0.50
	$\hat{\gamma}_1$	0.96	0.84	1.07
	γ_1^*	1.00	0.88	1.11
	$\hat{\gamma}_2$	0.91	0.93	1.04
	γ_2^*	1.00	0.95	1.02

Table 3.2: Comparison of simulated and least false values for $n = 10000, 50000$

n		SP	MAR	MNAR
10000	$\bar{R}$	0.50	0.50	0.50
	$E[R]$	0.50	0.50	0.50
	$\overline{Rx}$	0.01	-0.20	-0.10
	$E[Rx]$	0.00	-0.20	-0.11
	$\overline{RD}$	0.51	0.24	0.45
	$E[RD]$	0.50	0.25	0.45
	$\overline{RDx}$	0.50	0.31	0.42
	$E[RDx]$	0.50	0.30	0.39
	$\overline{Rx^2}$	0.50	0.51	0.52
	$E[Rx^2]$	0.50	0.50	0.50
	$\hat{\gamma}_1$	1.01	0.85	1.12
	γ_1^*	1.00	0.88	1.11
	$\hat{\gamma}_2$	1.00	0.94	1.02
	γ_2^*	1.00	0.95	1.02
50000	$\overline{R}$	0.50	0.50	0.49
	$E[R]$	0.50	0.50	0.50
	$\overline{Rx}$	0.00	-0.20	-0.11
	$E[Rx]$	0.00	-0.20	-0.11
	$\overline{RD}$	0.50	0.25	0.44
	$E[RD]$	0.50	0.25	0.45
	$\overline{RDx}$	0.51	0.30	0.39
	$E[RDx]$	0.50	0.30	0.39
	$\overline{Rx^2}$	0.50	0.50	0.50
	$E[Rx^2]$	0.50	0.50	0.50
	$\hat{\gamma}_1$	1.01	0.89	1.13
	γ_1^*	1.00	0.88	1.11
	$\hat{\gamma}_2$	1.01	0.95	1.03
	γ_2^*	1.00	0.95	1.02

3.4.2   The 95% reference interval for the expectations and least false β_3^*, β_4^*, γ_1^* and γ_2^*.

Here we will show that for each limiting value calculated by the expectation, the simulated values (SV) are within noise of the theoretical values (TV) for large sample sizes. We estimate the noise from the simulations; that is we get a confidence interval from the simulations and reassurance that the population values are within these. We simulate 100 samples of size 50000, then find the $\hat{\beta}$ for each sample then calculate the reference interval instead of standard confidence interval, so we will say: CI=$\bar{\hat{\beta}} \pm 2\,SD(\beta)$. We did this for the SP, MAR and MNAR dropout models, each with just one set of dropout parameters and all with the missingness percentage at about 50% dropout. The reason for that is if we used the standard confidence interval, which is CI=$\bar{\hat{\beta}} \pm 2\,\frac{SD(\beta)}{\sqrt{100}}$ as the CI for the true least false parameter, and if we took the number of repetitions up from 100 to many many more, it would have zero width. But the TV of course is an approximation, as we had to move from logit to probit form to get the skew normal results, it is not the exact value. We know it will not be exactly equal to the true least false parameter, and so for large enough number of repetitions will never be in the CI. Therefore, what we do is change from CI to reference interval using $\pm 2SD$. This should show that the sample values are within noise of the TV.

Table 3.3: The 95% reference interval for the expectations under SP using 100 samples of size 50000.

	TV	Mean	Lower Bound	Upper Bound
$E[R]$	0.5000	0.5030	0.4720	0.5340
$E[Rx]$	0.0000	-0.0212	-0.0657	0.0232
$E[RD]$	0.5000	0.5062	0.4392	0.5732
$E[RDx]$	0.5000	0.4698	0.3813	0.5583
$E[Rx^2]$	0.5000	0.5143	0.4471	0.5815

Table 3.4: The 95% reference interval for the expectations under MAR using 100 samples of size 50000.

	TV	Mean	Lower Bound	Upper Bound
$E[R]$	0.5000	0.4920	0.4610	0.5230
$E[Rx]$	-0.1960	-0.1972	-0.2392	-0.1551
$E[RD]$	0.2550	0.2473	0.1879	0.3068
$E[RDx]$	0.3040	0.2888	0.2202	0.3573
$E[Rx^2]$	0.5000	0.4992	0.4327	0.5657

Table 3.5: The 95% reference interval for the expectations under MNAR using 100 samples of size 50000.

	TV	Mean	Lower Bound	Upper Bound
$E[R]$	0.5000	0.4980	0.4670	0.5290
$E[Rx]$	-0.1083	-0.1368	-0.1832	-0.0903
$E[RD]$	0.4458	0.4368	0.3681	0.5056
$E[RDx]$	0.3917	0.4314	0.3462	0.5166
$E[Rx^2]$	0.5000	0.5796	0.4985	0.6608

Table 3.6: The 95% reference interval for β_3^* using 100 samples of size 50000.

	TV	SV	Lower	Upper
MAR	-1.1158	-1.1192	-1.1288	-1.1096
MNAR	-0.8864	-0.8816	-0.8911	-0.8721
SP	-1.0000	-0.9995	-1.0081	-0.9910

Table 3.7: The 95% reference interval for β_4^* using 100 samples of size 50000.

	TV	SV	Lower	Upper
MAR	-1.0454	-1.0462	-1.0791	-1.0133
MNAR	-0.9754	-0.9733	-1.0013	-0.9454
SP	-1.0000	-1.0013	-1.0338	-0.9689

Table 3.8: The 95% reference interval for γ_1^* using 100 samples of size 50000.

	TV	SV	Lower	Upper
MAR	0.8842	0.8790	0.8523	0.9057
MNAR	1.1136	1.1169	1.0859	1.1480
SP	1.0000	1.0043	0.9760	1.0327

Table 3.9: The 95% reference interval for γ_2^* using 100 samples of size 50000.

	TV	SV	Lower	Upper
MAR	0.9546	0.9546	0.9546	0.9546
MNAR	1.0246	1.0269	0.9970	1.0569
SP	1.0000	0.9994	0.9678	1.0311

3.4.3 The effect of dropout on the limiting values γ_1^* and γ_2^*

In Table 3.10, we show γ_1^* and γ_2^* by solving the equations given in Sections 3.3.5 and 3.3.6 for the given values of θ_0^{sp} and θ_1^{sp} under MAR, with the same parameters used to generate the logistic model as mentioned before. These values of θ all give a dropout rate of around 50%. In Table 3.11, we do the same for MNAR.

Table 3.10: Effect on dropout parameters of fitting the LI method under MAR, where $Bias(\gamma_1)$ and $Bias(\gamma_2)$ are the absolute bias in γ_1 and γ_2, respectively.

θ_0^M	θ_1^M	γ_1^*	γ_2^*	$Bias(\gamma_1)$	$Bias(\gamma_2)$
1	0.5	0.88	0.95	0.12	0.05
-1	-0.5	1.12	0.95	0.12	0.05
2	1	0.79	0.88	0.21	0.12
-2	-1	1.21	0.88	0.21	0.12

Table 3.11: Effect on dropout parameters of fitting the LI method under MNAR, where $Bias(\gamma_1)$ and $Bias(\gamma_2)$ are the absolute bias in γ_1 and γ_2, respectively.

θ_0^{MN}	θ_2^{MN}	γ_1^*	γ_2^*	$Bias(\gamma_1)$	$Bias(\gamma_2)$
0.5	0.5	1.11	1.02	0.11	0.02
-0.5	-0.5	0.89	1.02	0.11	0.02
1	1	1.21	1.07	0.21	0.07
-1	-1	0.79	1.07	0.21	0.07

Results in Table 3.10 and Table 3.11 are of main interest. Under MAR, Table 3.10 shows that the absolute bias in both γ_1^* and γ_2^* become larger as the absolute value of θ_1^M get bigger. For negative θ_1^M, we note that γ_1^* tends to be higher than its true value. The opposite happens for positive θ_1^M.

Turning to the next table, under MNAR, γ_1^* tends to be either more or less than its true value depending upon whether dropout is associated with large Y_2 (negative θ_2^{MN}), or small Y_2 (positive θ_2^{MN}). If dropout is associated with large Y_2, then the mean in the observed data will be too low, and the wrong missingness model does not compensate for this. The opposite is true if dropout is associated with small Y_2.

3.4.4   The effect of measurement errors of the random effect and covariate variances on the limiting values γ_1^* and γ_2^*

In the following tables we will study the effect of changing the variance of the random effect and measurement error on the limiting values at different dropout models. We take $\sigma_U^2 = \sigma_{\epsilon_1}^2 = \sigma_{\epsilon_2}^2$. Each time, we will keep the same dropout percentage at about 50%.

Table 3.12: The effect of the variance σ_U^2 on the limiting values γ_1^* and γ_2^*.

		SP	MAR	MNAR
$\sigma_U = \sqrt{0.5}$	γ_1^*	1.0000	0.8842	1.1136
	γ_2^*	1.0000	0.9546	1.0246
	$Bias(\gamma_1)$	0.0000	0.1158	0.1136
	$Bias(\gamma_2)$	0.0000	0.0454	0.0246
$\sigma_U = 1$	γ_1^*	1.0000	0.7774	1.2186
	γ_2^*	1.0000	0.9152	1.0457
	$Bias(\gamma_1)$	0.0000	0.2226	0.2186
	$Bias(\gamma_2)$	0.0000	0.0848	0.0457
$\sigma_U = \sqrt{3}$	γ_1^*	1.0000	0.4156	1.5752
	γ_2^*	1.0000	0.7992	1.1065
	$Bias(\gamma_1)$	0.0000	0.5844	0.5752
	$Bias(\gamma_2)$	0.0000	0.2008	0.1065
$\sigma_U = \sqrt{6}$	γ_1^*	1.0000	-0.0046	1.9916
	γ_2^*	1.0000	0.6947	1.1596
	$Bias(\gamma_1)$	0.0000	1.0046	0.9916
	$Bias(\gamma_2)$	0.0000	0.3053	0.1596
$\sigma_U = \sqrt{12}$	γ_1^*	1.0000	-0.6269	2.6114
	γ_2^*	1.0000	0.5873	1.2126
	$Bias(\gamma_1)$	0.0000	1.6269	1.6114
	$Bias(\gamma_2)$	0.0000	0.4127	0.2126

Changing the variance parameters and keeping the dropout percentage at about 50% has a remarkable effect on the bias. From Table 3.12, it is clear that as σ_U (and consequentially σ_{ϵ_1} and σ_{ϵ_2}) increases, the limiting values γ_1^* and γ_2^* go further from the true value ($\gamma_1^G = 1, \gamma_2^G = 1$) and hence the absolute bias increases, which means that the large error variances imply poor results. In short we can conclude that the more variability, the more bias. In Table 3.13 we turn to the effect of variance of the covariate, σ_x^2.

Table 3.13: The effect of the variance σ_x^2 on the limiting values γ_1^* and γ_2^*.

		SP	MAR	MNAR
$\sigma_x=\sqrt{0.5}$	γ_1^*	1.0000	0.8855	1.1131
	γ_2^*	1.0000	0.9521	1.0250
	$Bias(\gamma_1)$	0.0000	0.1145	0.1131
	$Bias(\gamma_2)$	0.0000	0.0479	0.0250
$\sigma_x=\sqrt{3}$	γ_1^*	1.0000	0.8832	1.1153
	γ_2^*	1.0000	0.9624	1.0233
	$Bias(\gamma_1)$	0.0000	0.1168	0.1153
	$Bias(\gamma_2)$	0.0000	0.0376	0.0233
$\sigma_x=\sqrt{9}$	γ_1^*	1.0000	0.8916	1.1170
	γ_2^*	1.0000	0.9752	1.0201
	$Bias(\gamma_1)$	0.0000	0.1084	0.1170
	$Bias(\gamma_2)$	0.0000	0.0248	0.0201
$\sigma_x=\sqrt{12}$	γ_1^*	1.0000	0.8981	1.1167
	γ_2^*	1.0000	0.9798	1.0184
	$Bias(\gamma_1)$	0.0000	0.1019	0.1167
	$Bias(\gamma_2)$	0.0000	0.0202	0.0184
$\sigma_x=\sqrt{16}$	γ_1^*	1.0000	0.9026	1.1160
	γ_2^*	1.0000	0.9822	1.0173
	$Bias(\gamma_1)$	0.0000	0.0974	0.1160
	$Bias(\gamma_2)$	0.0000	0.0178	0.0173
$\sigma_x=\sqrt{20}$	γ_1^*	1.0000	0.9078	1.1148
	γ_2^*	1.0000	0.9847	1.0160
	$Bias(\gamma_1)$	0.0000	0.0922	0.1148
	$Bias(\gamma_2)$	0.0000	0.0153	0.0160

This time changing σ_x has little effect on the bias.

- The variance effect at MAR dropout: As σ_x increases the limiting value γ_1^* gets slightly closer to the true value which implies smaller absolute bias.

- The variance effect at MNAR dropout: Here we note that the effect on γ_1^* differs from the effect on γ_2^*. Table 3.13 shows that as σ_x increases the limiting value γ_1^* goes slightly further from the true value and hence there is slightly bigger absolute bias. On the other hand, the limiting value γ_2^* gets slightly closer to the true value as σ_x increases or we can say that in this case the absolute bias get smaller. But in both cases the effect is small.

3.4.5   Contour plots

Here we show how the limiting values γ_1^* and γ_2^* change when we change the MAR or MNAR parameters. Figures 3.1 to 3.4 present contour plots of least false values γ_1^* and γ_2^* for a grid with $-1.5 < \theta_0 = \theta_0^{sp} < 1.5$ and $-1.5 < \theta_1 = \theta_1^{sp} < 1.5$. The range of $P(R = 1|Y_1, Y_2)$ over the grid is (0.2119, 0.7881) under MAR and it is (0.1987 ,0.8013) under MNAR. The legend on the right side of the plot represents the range of the parameter under study. For example in Figure 3.1, γ_1^* under MAR can take a value from about 0.7 to 1.3 for the θs changes in aforementioned grid. The white area in the plot corresponds to the correct value which is $\gamma_1^* = 1$.

In Figure 3.1 the biggest change is in the vertical direction, showing that θ_1 is most important. For example, at $\theta_1 = 0$ (which is equivalent to MCAR) then $\gamma_1^* = 1$ as expected. For positive θ_1, dropout is associated with small Y_1, so D tends to be low. Hence γ_1^* is lower than it should be. The opposite happens for a negative θ_1.

In Figure 3.2 we get negative bias as θ_1 moves away from zero in either direction. Such an attenuation of regression effect is common when there are errors in variables, see Carrol et al. [1995]. We speculate that a similar effect is in play here, for more details see Appendix A.6.

Turning to MNAR, Figures 3.3 and 3.4 show least false γ_1^* and γ_2^*. In Figure 3.3 as mentioned above for Figure 3.1, the biggest change here in Figure 3.3 is in the vertical direction, showing that θ_2 is most important. For example, at $\theta_2 = 0$ (which is equivalent to MCAR) then $\gamma_1^* = 1$ as expected. For negative θ_2, dropout is associated with large Y_2, so D tends to be low. Hence γ_1^* is lower than it should be. The opposite happens for a positive θ_2.

Figure 3.4 is similar to Figure 3.2, in that we get positive bias as θ_2 moves away from zero in either direction. This is the opposite to the MAR effect, because $D = Y_2 - Y_1$ and we expect selectivity in Y_2 to have the opposite effect to the same selectivity in Y_1 because of the different sign.

Figure 3.1: Contour plot of γ_1^* under MAR
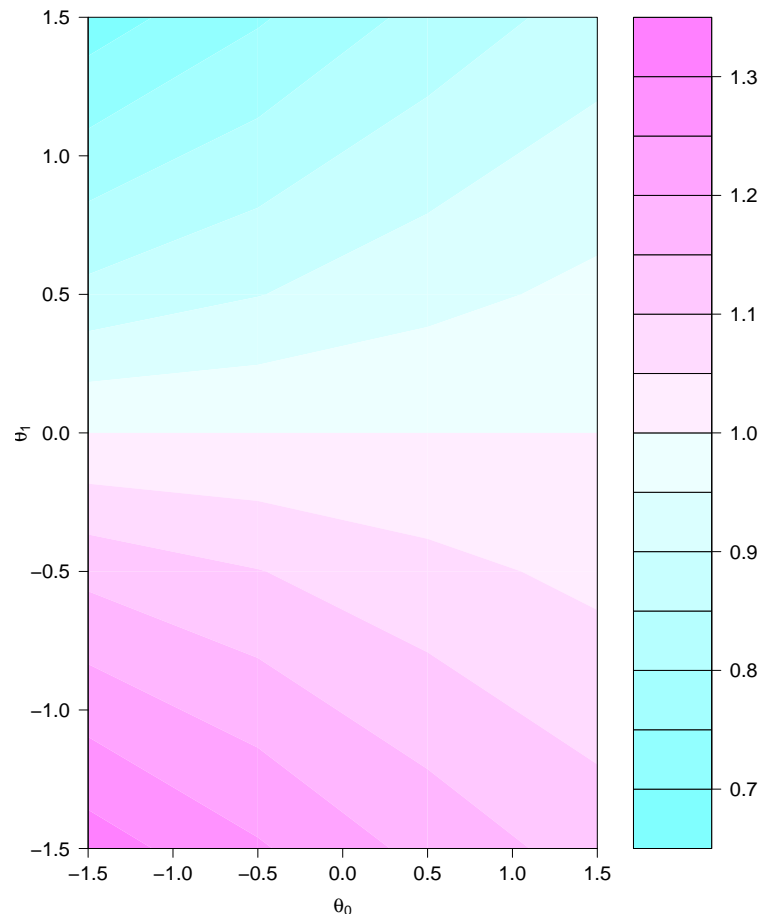
Figure 3.2: Contour plot of γ_2^* under MAR
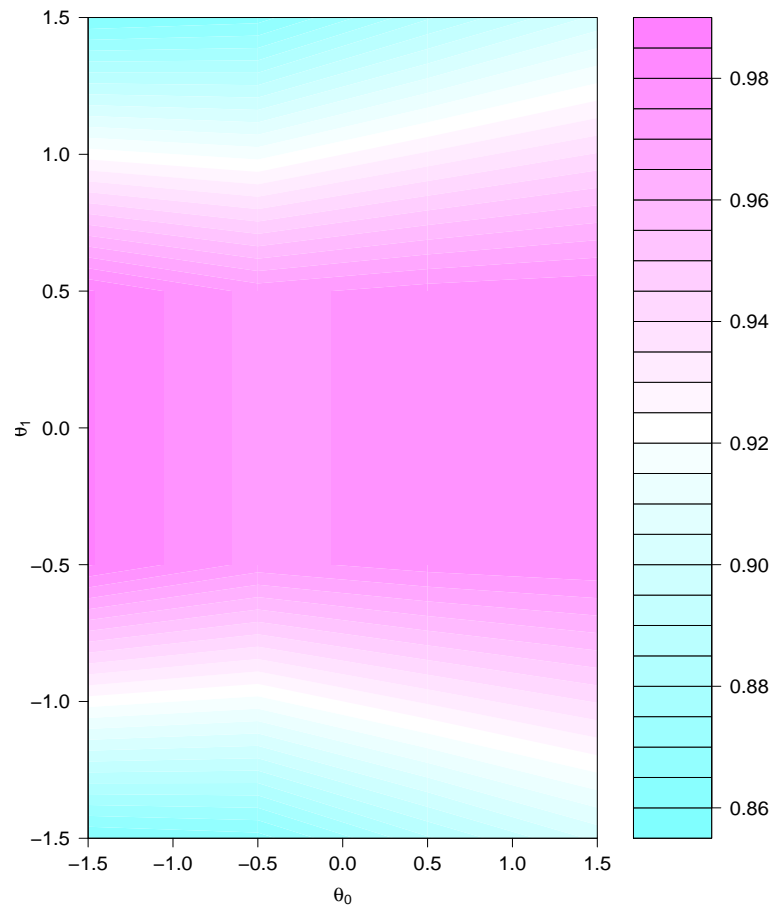
Figure 3.3: Contour plot of γ_1^* under MNAR
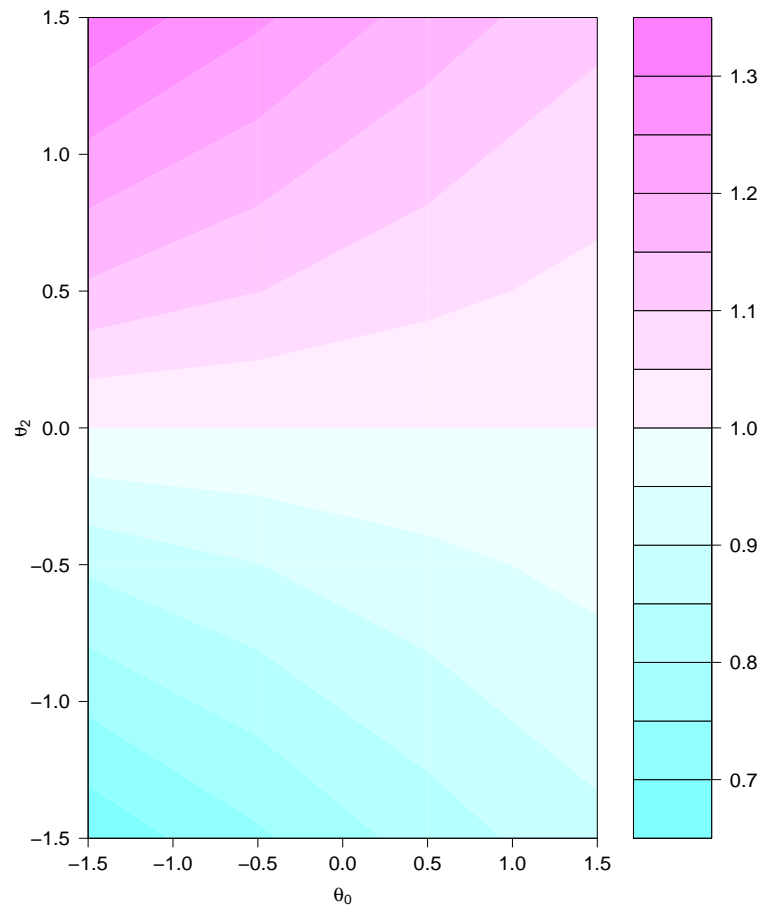
Figure 3.4: Contour plot of γ_2^* under MNAR
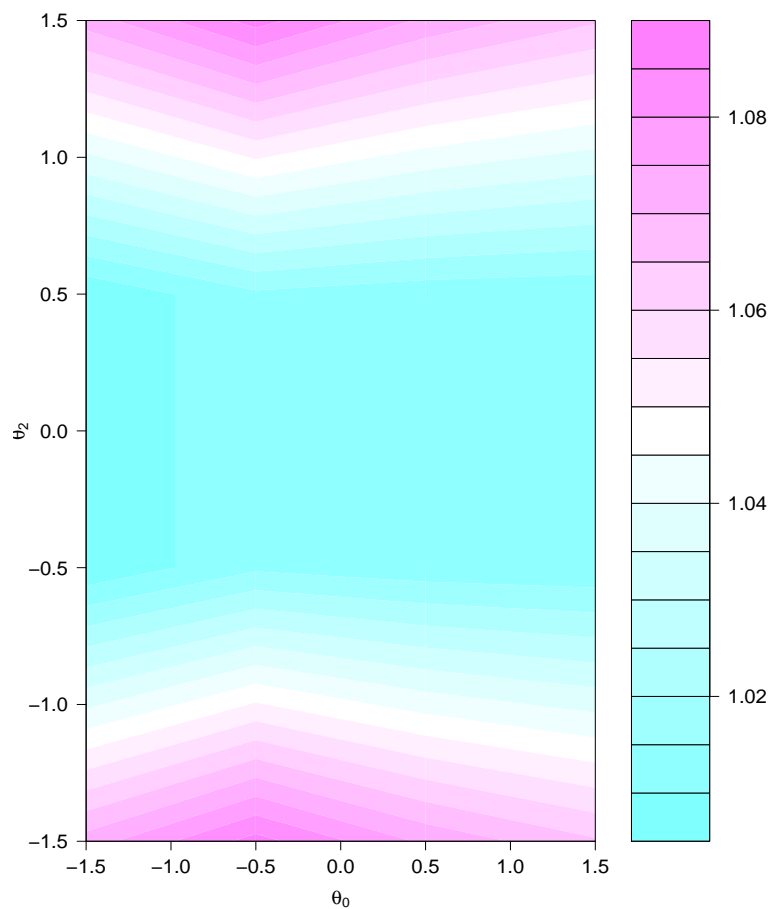
3.5   Discussion and Conclusion

In this chapter we calculated the expected value of terms in equations (3.15) and (3.14), then found the least false values of γ_1^* and γ_2^*. We compared these expected values under different dropout models to the generating values γ_1^G and γ_2^G. We calculated the bias in the expected parameters at each dropout model. It is clear now that the LI works under shared parameter, since $\gamma_2^* = \gamma_2^G$ and $\gamma_1^* = \gamma_1^G$ or in other words, the bias is zero. In contrast, the LI method does not work under MAR and MNAR as there is bias in both parameters. Results from the contour plots support our conclusions from Table 3.10 and Table 3.11.

Chapter 4

Performance of Inverse Probability Weighting (IPW) Under Shared Parameter, MAR and MNAR Dropout

4.1 Introduction

In this chapter we aim to investigate how the IPW method performs under the shared parameter, MAR and MNAR dropout mechanisms. The derivation and illustration of least false values are made under SP and MNAR dropout. Under MAR we show that the method gives consistent estimates. We make a comparison of least-false values with simulation results for selected parameter combinations to check our calculations.

Aalen and Gunnes [2010] compared the linear increment method with inverse probability weighting for complete data, and found that the latter method is more general, while the increment method depends on the assumptions of model linearity. However, they argued that in a missing data case the linear increments model should usually considered to be a preferred option over using the inverse probability weighting methods. Matthews et al. [2012] show that the inverse probability technique has a higher standard error compared to other techniques as we mentioned in Chapter 2. In general, Diggle et al. [2007] found that the inverse probability weighting method is inefficient, however they did not explain why or how. We have looked at the estimates and the standard errors in Chapter 2, and we found similar results.

For references regarding the IPW method, see Horvitz and Thompson [1952], Robins et al. [1995] and Hernan et al. [2006]. A comparison between the present method and the multiple imputation method was conducted by Carpenter et al. [2006]. Ho et al. [2012] used the inverse probability weighting (IPW) method, which was proposed by Robins and

colleagues (Robins et al. [1995] and Rotnitzky et al. [1998]). Assuming the data are missing at random (MAR), but the fact that the data are missing not at random (MNAR), Ho et al. [2012] investigated the effect of treatment effect estimation under a crossover design. In this case, the MAR analysis will be biased. They found that the bias is affected by the parameters, and it could be lower for some chosen sets than for others.

An investigation into the effect of choosing the parameters on the bias was also shown by Matthews et al. [2012]. We discuss this in our simulation in Section 4.6.

Section 4.2 introduces the two timepoints model as a special case. In Section 4.3 we investigate the performance of the IPW method under the MAR model. The performance of the IPW method under MNAR and SP models is shown in Sections 4.4 and 4.5, respectively. They include brief details of how to use the extended skewed normal density to find the expectation needed as in Chapter 3. Section 4.6 presents the results from analyses of different simulation studies. Finally, in Section 4.7 we discuss all of our results.

4.2 Two Timepoints As a Special Case: Generating Model

4.2.1 Longitudinal data model

For this chapter we will assume that we have n individuals and two repeated measurements or in other words two timepoints. Our model is the same as the model used in the previous chapter.

4.2.2 Estimation of the parameters

According to Ho et al. [2012] and the recap of the inverse probability weighting approach which is illustrated in Chapter 2, the closed form of the inverse probability estimate of β is:

$$\hat{\beta} = \left(\sum \frac{R_i X_i^T X_i}{\pi_i}\right)^{-1} \left(\sum \frac{R_i X_i^T Y_i}{\pi_i} - \sum R_i \frac{1 - \pi_i}{\pi_i} \phi_1(Y_1, x) + \sum (1 - R_i)\phi_1(Y_1, x)\right),$$
(4.1)

where $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4)^T$, $\pi_i = \text{expit}\{\theta_0^* + \theta_1^* Y_{1i}\}$. Here π_i is the MAR dropout probability, possibly estimated. The function ϕ_1 is arbitrary and is investigator chosen. In our work we can simply ignore ϕ_1 and rewrite $\hat{\beta}$ as:

$$\hat{\beta} = \left(\sum \frac{R_i X_i^T X_i}{\pi_i}\right)^{-1} \sum \frac{R_i X_i^T Y_i}{\pi_i}.$$
(4.2)

We can split $\hat{\beta}$ into:

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \left[\sum \frac{R_i}{\pi_i} \begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix}\right]^{-1} \left(\sum \frac{R_i}{\pi_i} \begin{pmatrix} Y_{1i} \\ Y_{1i} x_i \end{pmatrix}\right),$$
(4.3)

and

$$\begin{pmatrix} \hat{\beta}_3 \\ \hat{\beta}_4 \end{pmatrix} = \left[\sum \frac{R_i}{\pi_i} \begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix} \right]^{-1} \left(\sum \frac{R_i}{\pi_i} \begin{pmatrix} Y_{2i} \\ Y_{2i}x_i \end{pmatrix} \right). \tag{4.4}$$

Since the dropout occurs at the second timepoint, we can concentrate on $(\hat{\beta}_3, \hat{\beta}_4)$, and thus the coefficients in the model for Y_2 are of most interest. We will now work on the components of equation (4.4).

Clearly

$$\sum \frac{R_i}{\pi_i} \begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix} = \begin{pmatrix} \sum \frac{R_i}{\pi_i} & \sum \frac{R_i}{\pi_i}x_i \\ \sum \frac{R_i}{\pi_i}x_i & \sum \frac{R_i}{\pi_i}x_i^2 \end{pmatrix}$$

and also

$$\left[\sum \frac{R_i}{\pi_i} \begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix} \right]^{-1} = \frac{1}{\left(\sum \frac{R_i}{\pi_i} \right) \left(\sum \frac{R_i x_i^2}{\pi_i} \right) - \left(\sum \frac{R_i x_i}{\pi_i} \right)^2} \begin{pmatrix} \sum \frac{R_i}{\pi_i}x_i^2 & -\sum \frac{R_i}{\pi_i}x_i \\ -\sum \frac{R_i}{\pi_i}x_i & \sum \frac{R_i}{\pi_i} \end{pmatrix}.$$

Dividing top and bottom by n allows us to change the sums to means, which is helpful for later use of the laws of large numbers (LLN). We will study what happens to the estimates under the three different dropout models, as n increases.

Let $n \to \infty$. We use a superscript $*$ to imply the limiting value. As sample size increases, the sample means converges to the corresponding expected values. This implies that:

$$\begin{pmatrix} \hat{\beta}_3 \\ \hat{\beta}_4 \end{pmatrix} \to \begin{pmatrix} \beta_3^* \\ \beta_4^* \end{pmatrix} = E\left[\frac{R}{\pi} \begin{pmatrix} 1 & x \\ x & x^2 \end{pmatrix} \right]^{-1} E\left[\frac{R}{\pi} \begin{pmatrix} Y_2 \\ Y_2 x \end{pmatrix} \right]. \tag{4.5}$$

Taking the expectation to be inside the array:

$$\begin{pmatrix} \beta_3^* \\ \beta_4^* \end{pmatrix} = \begin{pmatrix} E[\frac{R}{\pi}] & E[\frac{Rx}{\pi}] \\ E[\frac{Rx}{\pi}] & E[\frac{Rx^2}{\pi}] \end{pmatrix}^{-1} \begin{pmatrix} E[\frac{RY_2}{\pi}] \\ E[\frac{RY_2x}{\pi}] \end{pmatrix}. \tag{4.6}$$

Hence to find the limiting values β_3^* and β_4^*, we need to calculate the expectations $E[R/\pi]$, $E[Rx/\pi]$, $E[Rx^2/\pi]$, $E[RY_2/\pi]$ and $E[RY_2x/\pi]$.

In the above we have assumed π is known. We will assume that if we have to estimate π then we will make the assumption, perhaps false, of MAR dropout. We need first to investigate what happens to our estimates as n increases.

Let $\hat{\pi} = e^{\hat{\theta}_0 + \hat{\theta}_1 Y_{1i}} / \{1 + e^{\hat{\theta}_0 + \hat{\theta}_1 Y_{1i}}\}$. The likelihood and score are

$$\ell = \sum (R_i \log \hat{\pi}_i + (1 - R_i) \log (1 - \hat{\pi}_i))$$

$$\frac{\partial \ell}{\partial \theta} = \sum \left(\frac{R_i}{\hat{\pi}_i} - \frac{1 - R_i}{1 - \hat{\pi}_i} \right) \begin{pmatrix} 1 \\ Y_{1i} \end{pmatrix} \hat{\pi}_i (1 - \hat{\pi}_i)$$

$$= \sum (R_i (1 - \hat{\pi}_i) - \hat{\pi}_i (1 - R_i)) \begin{pmatrix} 1 \\ Y_{1i} \end{pmatrix}.$$

Rearranging, we get the estimating equation:

$$\sum (R_i - \hat{\pi}_i) \begin{pmatrix} 1 \\ Y_{1i} \end{pmatrix} = 0. \tag{4.7}$$

Again, we divide the right and left sides of equation (4.7) by n, and note that since the right side is zero it stays zero. Using the law of large numbers (LLN), let $n \to \infty$, which implies $\hat{\theta} \to \theta^*$, where the parameter of interest is made up of the least false values θ_0^* and θ_1^* or what is sometimes called "the best approximating value", see Claeskens and Hjort [2008]. We can find them by applying the LLN, which allows us to change the sums to expectations in equation (4.7), and this gives:

$$E \left[(R - \pi) \begin{pmatrix} 1 \\ Y_1 \end{pmatrix} \right] = 0. \tag{4.8}$$

Equation (4.8) can be split into two equations:

$$E[R] - E[\text{expit} \{\theta_0^* + \theta_1^* Y_1\}] = 0. \tag{4.9}$$

$$E[Y_1 R] - E[Y_1 \text{ expit} \{\theta_0^* + \theta_1^* Y_1\}] = 0. \tag{4.10}$$

We will study what happens to the estimates under the three different dropout models, as n increases. In the next sections we will try to find the components of equations (4.9) and (4.10) under different dropout models. We assume the fitted model is MAR: $P(R = 1|Y_1) = e^{\theta_0^M + \theta_1^M Y_1} / \{1 + e^{\theta_0^M + \theta_1^M Y_1}\}$, but the truth is either the MNAR model $P(R = 1|Y_1, Y_2) = e^{\theta_0^{MN} + \theta_2^{MN} Y_2} / \{1 + e^{\theta_0^{MN} + \theta_2^{MN} Y_2}\}$, or it might be the SP model $P(R = 1|U) = e^{\theta_0^s + \theta_1^s U} / \{1 + e^{\theta_0^s + \theta_1^s U}\}$.

4.3   Performance under MAR

If the true dropout model is MAR and we fit the same model by maximum likelihood, then clearly the estimates are consistent and $\theta_0^* = \theta_0^M$ and $\theta_1^* = \theta_1^M$. Also $E_{R|Y_1, Y_2, x}[R] = \pi$, where $\pi = E[\text{expit}(\theta_0^M + \theta_1^M Y_1)]$. Turning to the regression parameters β, from the results found in Chapter 2 (Tables 2.3 and 2.4), we know that the IPW method provides consistent

estimates under MAR. We will demonstrate this for completeness. We use subscripts to indicate which variables we are considering when taking expectations, unless it is perfectly clear. To see what $\hat{\beta}$ converges to, in this case as MAR is correct model, then we can work on equation (4.5):

$$\begin{pmatrix} \beta_3^* \\ \beta_4^* \end{pmatrix} = E\left[\frac{R}{\pi} \begin{pmatrix} 1 & x \\ x & x^2 \end{pmatrix} \right]^{-1} E\left[\frac{R}{\pi} \begin{pmatrix} Y_2 \\ Y_2 x \end{pmatrix} \right]. \tag{4.11}$$

We can take each part separately:

$$\begin{aligned} E\left[\frac{R}{\pi} \begin{pmatrix} Y_2 \\ Y_2 x \end{pmatrix} \right] &= E_{x,Y_1,Y_2,R}\left[\frac{R}{\pi} \begin{pmatrix} Y_2 \\ Y_2 x \end{pmatrix} \right] \\ &= E_{x,Y_1,Y_2}\left[E_{R|Y_1,Y_2,x}\left[\frac{R}{\pi} \begin{pmatrix} Y_2 \\ Y_2 x \end{pmatrix} \right] \right], \end{aligned} \tag{4.12}$$

but we can say $E_{R|Y_1,Y_2,x}\left[R/\pi \right] = 1$. Then,

$$\begin{aligned} E\left[\frac{R}{\pi} \begin{pmatrix} Y_2 \\ Y_2 x \end{pmatrix} \right] &= E_{x,Y_1,Y_2}\left[\begin{pmatrix} Y_2 \\ Y_2 x \end{pmatrix} \right] \\ &= \begin{pmatrix} \beta_3 + \beta_4 E[x] \\ \beta_3 E[x] + \beta_4 E[x^2] \end{pmatrix}. \end{aligned} \tag{4.13}$$

Note that for simplicity we write $\begin{pmatrix} \beta_3 \\ \beta_4 \end{pmatrix}$ instead of $\begin{pmatrix} \beta_3^G \\ \beta_4^G \end{pmatrix}$.

Similarly,

$$\begin{aligned} E\left[\frac{R}{\pi} \begin{pmatrix} 1 & x \\ x & x^2 \end{pmatrix} \right]^{-1} &= \begin{pmatrix} 1 & E(x) \\ E(x) & E(x^2) \end{pmatrix}^{-1} \\ &= \frac{1}{(E(x^2) - E(x)^2)} \begin{pmatrix} E(x^2) & -E(x) \\ -E(x) & 1 \end{pmatrix}. \end{aligned} \tag{4.14}$$

Then, from equations (4.13) and (4.14):

$$\begin{aligned} \begin{pmatrix} \beta_3^* \\ \beta_4^* \end{pmatrix} &= \frac{1}{(E(x^2) - E(x)^2)} \begin{pmatrix} \beta_3(E(x^2) - E(x)^2) \\ \beta_4(E(x^2) - E(x)^2) \end{pmatrix} \\ &= \begin{pmatrix} \beta_3 \\ \beta_4 \end{pmatrix} \end{aligned}$$

as expected.

4.4 Performance under Shared Parameter Dropout

Under SP as the true model and MAR as the assumed model, we will estimate the least false values θ_0^* and θ_1^* using equations (4.9) and (4.10). Recall:

$$E[R] - E[\text{expit } \{\theta_0^* + \theta_1^* Y_1\}] = 0.$$
$$E[Y_1 R] - E[Y_1 \text{ expit } \{\theta_0^* + \theta_1^* Y_1\}] = 0.$$

We will use an approach similar to that used in Chapter 3 based on the extended skew normal distribution (ESN).

4.4.1 Calculating the components of equation (4.9)

We start with the first part of equation (4.9). Since under SP, R depends only on U, then

$$
\begin{aligned}
E[R] &= \int P(R = 1|U) f(U) \, \mathrm{d}U \\
&= \int \text{expit } \{\theta_0 + \theta_1 U\} \phi(U; 0, \sigma_U^2) \, \mathrm{d}U \\
&\approx \int \Phi[c(\theta_0 + \theta_1 U)] \phi(U; 0, \sigma_U^2) \, \mathrm{d}U & (4.15) \\
&= \Phi\left[\frac{c\,\theta_0}{\sqrt{1 + (c\,\theta_1 \sigma_U)^2}}\right]. & (4.16)
\end{aligned}
$$

Note that in equation (4.15) we used the approximation of the expit to the cumulative normal, which we defined in Chapter 3, see Johnson et al. [1995]:

$$\text{expit}(z) \approx \Phi(c\,z), \, c = \frac{16\sqrt{3}}{15\pi}. \tag{4.17}$$

Also, we moved to equation (4.16) by using ESN; because the integral is now equivalent to the left side of formula (A.2) in Appendix A.1. For more details refer to Appendix A.1.

For the second part of equation (4.9) we have

$$E[\text{ expit } \{\theta_0^* + \theta_1^* Y_1\}] = E[\text{ expit } \{\theta_0^* + \theta_1^*(\beta_1 + \beta_2 x + U + \epsilon_1)\}] = E[R^*], \text{ say.}$$

Thus R^* depends on x, U and ϵ_1. We need to calculate this expectation in two stages.

First, we calculate the expectation conditional upon x:

$$E[R^*|x] = \int \text{ expit } \{\theta_0^* + \theta_1^*(\beta_1 + \beta_2 x) + \theta_1^*(U + \epsilon_1)\} f(U, \epsilon_1) \, \mathrm{d}U \, \mathrm{d}\epsilon_1.$$

Let $U + \epsilon_1 = \xi_1$, thus $\xi_1 \sim N(0, \sigma_{\xi_1}^2)$, where $\sigma_{\xi_1}^2 = \sigma_U^2 + \sigma_{\epsilon_1}^2$. So $\hspace{2cm}$ (4.18)

$$
\begin{aligned}
E[R^*|x] &\approx \int \Phi[c\,(\theta_0^* + \theta_1^*(\beta_1 + \beta_2 x)) + c\,\theta_1^* \xi_1]\phi(\xi_1; 0, \sigma_{\xi_1}^2)\,\mathrm{d}\xi_1 \qquad (4.19)\\
&= \Phi\left[\frac{c\,(\theta_0^* + \theta_1^*(\beta_1 + \beta_2 x))}{\sqrt{1 + (c\,\theta_1^* \sigma_{\xi_1})^2}}\right]\\
&= \Phi[A_1 + A_2\,x],
\end{aligned}
$$

where $A_1 = \frac{c\,(\theta_0^* + \theta_1^* \beta_1)}{\sqrt{1 + (c\,\theta_1^* \sigma_{\xi_1})^2}}$ and $A_2 = \frac{c\,\theta_1^* \beta_2}{\sqrt{1 + (c\,\theta_1^* \sigma_{\xi_1})^2}}$. Note that to calculate the integration in equation (4.19) we again used the ESN formula (A.2) in Appendix A.1 .

Second, we remove the conditioning on x to get

$$
\begin{aligned}
E[R^*] &= E[E[R^*|x]] \approx E\left[\Phi[A_1 + A_2\,x]\right]\\
&= \int \Phi[A_1 + A_2\,x]\phi(x; 0, \sigma_x^2)\,\mathrm{d}x\\
&= \Phi\left[\frac{A_1}{\sqrt{1 + (A_2\sigma_x)^2}}\right]. \qquad (4.20)
\end{aligned}
$$

Then from equation (4.16) and (4.20) we can rewrite equation (4.9) to be

$$
E[R] - E[R^*] \approx \Phi\left[\frac{c\,\theta_0}{\sqrt{1 + (c\,\theta_1\sigma_U)^2}}\right] - \Phi\left[\frac{A_1}{\sqrt{1 + (A_2\sigma_x)^2}}\right] = 0. \qquad (4.21)
$$

4.4.2 Calculating the components of equation (4.10)

In the following we will work on equation (4.10). We start with the first part:

$$
\begin{aligned}
E[Y_1 R] &= E[Y_1 \,\mathrm{expit}\,\{\theta_0 + \theta_2 U\}]\\
&= E[(\beta_1 + \beta_2\,x + U + \epsilon_1)\,\mathrm{expit}\,\{\theta_0 + \theta_1 U\}]\\
&= E[\beta_1\,\mathrm{expit}\,\{\theta_0 + \theta_1 U\}] + E[\beta_2\,x\,\mathrm{expit}\,\{\theta_0 + \theta_1 U\}]\\
&\quad + E[U\,\mathrm{expit}\,\{\theta_0 + \theta_1 U\}] + E[\epsilon_1\,\mathrm{expit}\,\{\theta_0 + \theta_1 U\}].
\end{aligned}
$$

We have $E[\beta_2\,x\,\mathrm{expit}\,\{\theta_0 + \theta_1 U\}] = E[\epsilon_1\,\mathrm{expit}\,\{\theta_0 + \theta_1 U\}] = 0$, knowing that x and ϵ_1 are independent of U and have zero means. Thus,

$$
\begin{aligned}
E[Y_1 R] &= E[\beta_1\,\mathrm{expit}\,\{\theta_0 + \theta_1 U\}] + E[U\,\mathrm{expit}\,\{\theta_0 + \theta_1 U\}]\\
&= D_1 + D_2, \text{ say.} \qquad (4.22)
\end{aligned}
$$

We will work on D_1 first and then D_2. We have:

$$
\begin{aligned}
D_1 &= E[\beta_1 \operatorname{expit}\{\theta_0 + \theta_1 U\}] \\
&\approx E[\beta_1 \Phi[c(\theta_0 + \theta_1 U)]] \\
&= \beta_1 \int \Phi[c(\theta_0 + \theta_1 U)]\, \phi(U; 0, \sigma_U^2)\, \mathrm{d}U \qquad (4.23) \\
&= \beta_1 \Phi\left[\frac{c\,\theta_0}{\sqrt{1 + (c\,\theta_1\,\sigma_U)^2}}\right]. \qquad (4.24)
\end{aligned}
$$

And

$$
\begin{aligned}
D_2 &= E[U \operatorname{expit}\{\theta_0 + \theta_1 U\}] \\
&\approx E[U\Phi[c(\theta_0 + \theta_1 U)]] \\
&= \int U\Phi[c(\theta_0 + \theta_1 U)]\, \phi(U; 0, \sigma_U^2)\, \mathrm{d}U \qquad (4.25) \\
&= \frac{c\,\sigma_U^2\,\theta_1}{\sqrt{1 + (c\,\theta_1\,\sigma_U)^2}}\phi(\bar{\nu}_1), \qquad (4.26)
\end{aligned}
$$

where $\bar{\nu}_1 = \frac{c\,\theta_0}{\sqrt{1+(c\,\theta_1\,\sigma_U)^2}}$. Here for D_1, we used formula (A.2) to find the integration in equation (4.23), while for D_2 we used formula (A.7) to find equation (4.25). These formulas can be found in Appendix A.1 .

Subsequently, from equations (4.24) and (4.26) we can rewrite equation (4.22) as:

$$
E[Y_1 R] \approx \beta_1 \Phi\left[\frac{c\,\theta_0}{\sqrt{1 + (c\,\theta_1\,\sigma_U)^2}}\right] + \frac{c\,\sigma_U^2\,\theta_1}{\sqrt{1 + (c\,\theta_1\,\sigma_U)^2}}\phi(\bar{\nu}_1). \qquad (4.27)
$$

Now we will work on the second part of equation (4.10),

$$
E[Y_1 \operatorname{expit}\{\theta_0^* + \theta_1^* Y_1\}] = E[(\beta_1 + \beta_2\, x + U + \epsilon_1) \operatorname{expit}\{\theta_0^* + \theta_1^*(\beta_1 + \beta_2\, x + U + \epsilon_1)\}].
$$

Recall $\xi_1 = \epsilon_1 + U$ as defined in equation (4.18), then

$$
\begin{aligned}
E[Y_1 \operatorname{expit}\{\theta_0^* + \theta_1^* Y_1\}] &= E[\beta_1 \operatorname{expit}\{\theta_0^* + \theta_1^*(\beta_1 + \beta_2\, x + \xi_1)\}] \\
&+ E[\beta_2\, x \operatorname{expit}\{\theta_0^* + \theta_1^*(\beta_1 + \beta_2\, x + \xi_1)\}] \\
&+ E[\xi_1 \operatorname{expit}\{\theta_0^* + \theta_1^*(\beta_1 + \beta_2\, x + \xi_1)\}] \\
&= D_3 + D_4 + D_5. \qquad (4.28)
\end{aligned}
$$

In the following we will work on equation (4.28). We will start with D_3[1]:

$$
\begin{aligned}
D_3 &\approx \beta_1 E[\Phi[c\,(\theta_0^* + \theta_1^*(\beta_1 + \beta_2\,x + \xi_1))]]] \\
&= \beta_1 E[\Phi[k_1 + k_2\,x + k_3\,\xi_1]], \quad \text{say} \\
&= \beta_1 \int\int \Phi[k_1 + k_2\,x + k_3\,\xi_1] f(x,\xi_1)\,\mathrm{d}x\,\mathrm{d}\xi_1.
\end{aligned}
$$

Under SP, x and ξ_1 are independent, which allows us to write the joint density function $f(x,\xi_1) = f(x)f(\xi_1)$, and, as defined earlier, x and ξ_1 are normally distributed, so we can use $\phi(.;\mu,\sigma^2)$ instead of $f(.)$. Hence the above integration will be

$$
\begin{aligned}
D_3 &\approx \beta_1 \int\int \Phi[k_1 + k_2\,x + k_3\,\xi_1] f(x)f(\xi_1)\,\mathrm{d}x\,\mathrm{d}\xi_1 \\
&= \beta_1 \int\int \Phi[k_1 + k_2\,x + k_3\,\xi_1]\phi(x;0,\sigma_x^2)\phi(\xi_1;0,\sigma_{\xi_1}^2)\,\mathrm{d}x\,\mathrm{d}\xi_1 \\
&= \beta_1 \int\left(\int \Phi[k_1 + k_2\,x + k_3\,\xi_1]\phi(x;0,\sigma_x)\,\mathrm{d}x\right)\phi(\xi_1;0,\sigma_{\xi_1}^2)\mathrm{d}\xi_1 \\
&= \beta_1 \int\left(\int (\Phi[k_2\,x + k_4]\phi(x;0,\sigma_x^2)\,\mathrm{d}x\right)\phi(\xi_1;0,\sigma_{\xi_1}^2)\mathrm{d}\xi_1.
\end{aligned}
$$

We will do the inner integration first using the ESN formula (A.2) in Appendix A.1 , thus

$$
\begin{aligned}
D_3 &\approx \beta_1 \int\left(\Phi\left[\frac{k_4}{\sqrt{1 + (k_2\sigma_x)^2}}\right]\right)\phi(\xi_1;0,\sigma_{\xi_1}^2)\mathrm{d}\xi_1 \\
&= \beta_1 \int\left(\Phi\left[\frac{k_1 + k_3\xi_1}{\sqrt{1 + (k_2\sigma_x)^2}}\right]\right)\phi(\xi_1;0,\sigma_{\xi_1}^2)\mathrm{d}\xi_1 \\
&= \beta_1 \int \Phi\left[A_3 + A_4\xi_1\right]\phi(\xi_1;0,\sigma_{\xi_1}^2)\mathrm{d}\xi_1.
\end{aligned}
$$

Again, this integration can be calculated using the same formula. The final result is:

$$
D_3 \approx \beta_1\,\Phi\left[\frac{A_3}{\sqrt{1 + (A_4\sigma_{\xi_1})^2}}\right], \tag{4.29}
$$

where $A_3 = k_1/\{\sqrt{1 + (k_2\sigma_x)^2}\}$, $A_4 = k_3/\{\sqrt{1 + (k_2\sigma_x)^2}\}$, and $k_1 = c(\theta_0^* + \theta_1^*\beta_1)$, $k_2 = c(\theta_1^*\beta_2)$, $k_3 = c\,\theta_1^*$, $k_4 = k_1 + k_3\xi_1$.
For D_4:

$$
\begin{aligned}
D_4 &\approx \beta_2\,E[x\,\Phi[c\,(\theta_0^* + \theta_1^*(\beta_1 + \beta_2\,x)) + c\,\theta_1^*\xi_1] \\
&= \beta_2 E[x\,\Phi[k_1 + k_2\,x + k_3\,\xi_1]] \\
&= \beta_2 \int\int x\,\Phi[k_1 + k_2\,x + k_3\,\xi_1] f(x,\xi_1)\,\mathrm{d}x\,\mathrm{d}\xi_1.
\end{aligned}
$$

But under SP, x and ξ_1 are independent, which again allows us to write the joint function

---

$f(x, \xi_1) = f(x)f(\xi_1)$, and, as previously, x and ξ_1 are normally distributed, so we can use $\phi(.; \mu, \sigma^2)$ instead of $f(.)$. Hence the above will be

$$
\begin{aligned}
D_4 &\approx \beta_2 \int\int x\,\Phi[k_1 + k_2\,x + k_3\,\xi_1]f(x)f(\xi_1)\,\mathrm{d}x\,\mathrm{d}\xi_1 \\
&= \beta_2 \int\int x\,\Phi[k_1 + k_2\,x + k_3\,\xi_1]\phi(x; 0, \sigma_x^2)\phi(\xi_1; 0, \sigma_{\xi_1}^2)\,\mathrm{d}x\,\mathrm{d}\xi_1 \\
&= \beta_2 \int x \left(\int (\Phi[k_5 + k_3\,\xi_1]\phi(\xi_1; 0, \sigma_{\xi_1}^2)\,\mathrm{d}\xi_1\right)\phi(x; 0, \sigma_x^2)\mathrm{d}x,
\end{aligned}
$$

where $k_5 = k_1 + k_2\,x$. We will do the inner integration first using the ESN formula (A.2) in Appendix A.1 , thus

$$
\begin{aligned}
D_4 &\approx \beta_2 \int x \left(\Phi\left[\frac{k_5}{\sqrt{1 + (k_3\sigma_{\xi_1})^2}}\right]\right)\phi(x; 0, \sigma_x^2)\mathrm{d}x \\
&= \beta_2 \int x\,\Phi\left[\frac{k_1 + k_2\,x}{\sqrt{1 + (k_3\sigma_{\xi_1})^2}}\right]\phi(x; 0, \sigma_x^2)\mathrm{d}x \\
&= \beta_2 \int x\,\Phi[A_5 + A_6\,x]\,\phi(x; 0, \sigma_x^2)\mathrm{d}x.
\end{aligned}
$$

Using formula (A.7) in Appendix A.1 to find the integral, we get

$$
D_4 \approx \beta_2 \frac{A_6\,\sigma_x^2}{k_6}\phi(\bar{\nu}_2), \tag{4.30}
$$

where $A_5 = \frac{k_1}{\sqrt{1+(k_3\sigma_{\xi_1})^2}}$, $A_6 = \frac{k_2}{\sqrt{1+(k_3\sigma_{\xi_1})^2}}$, $k_6 = \sqrt{1 + (A_6\sigma_x)^2}$, and $\bar{\nu}_2 = \frac{A_5}{k_6}$.

Similarly,

$$
\begin{aligned}
D_5 &\approx E[\xi_1\Phi[c\,(\theta_0^* + \theta_1^*(\beta_1 + \beta_2\,x)) + c\,\theta_1^*\xi_1]] \\
&= E[\xi_1\,\Phi[k_1 + k_2\,x + k_3\,\xi_1]] \\
&= \int\int \xi_1\,\Phi[k_1 + k_2\,x + k_3\,\xi_1]\phi(x; 0, \sigma_x^2)\phi(\xi_1; 0, \sigma_{\xi_1}^2)\,\mathrm{d}x\,\mathrm{d}\xi_1 \\
&= \int \xi_1 \left(\int \Phi[k_1 + k_2\,x + k_3\,\xi_1]\phi(x; 0, \sigma_x^2)\,\mathrm{d}x\right)\phi(\xi_1; 0, \sigma_{\xi_1}^2)\mathrm{d}\xi_1 \\
&= \int \xi_1 \left(\int (\Phi[k_2\,x + k_4]\phi(x; 0, \sigma_x^2)\,\mathrm{d}x\right)\phi(\xi_1; 0, \sigma_{\xi_1}^2)\mathrm{d}\xi_1.
\end{aligned}
$$

We will do the inner integration first using the ESN formula (A.2) in Appendix A.1, thus

$$
\begin{aligned}
D_5 &\approx \int \xi_1 \left(\Phi \left[\frac{k_4}{\sqrt{1 + (k_2 \sigma_x)^2}} \right] \right) \phi(\xi_1; 0, \sigma_{\xi_1}^2) \mathrm{d}\xi_1 \\
&= \int \xi_1 \, \Phi \left[\frac{k_1 + k_3 \, \xi_1}{\sqrt{1 + (k_2 \sigma_x)^2}} \right] \phi(\xi_1; 0, \sigma_{\xi_1}^2) \mathrm{d}\xi_1 \\
&= \int \xi_1 \, \Phi \left[A_3 + A_4 \xi_1 \right] \phi(\xi_1; 0, \sigma_{\xi_1}^2) \mathrm{d}\xi_1 .
\end{aligned}
$$

Using formula (A.7) in Appendix A.1 to find the integral,

$$
D_5 \quad \approx \quad \frac{A_4 \, \sigma_{\xi_1}^2}{k_7} \phi(\bar{\nu}_3), \tag{4.31}
$$

where $k_7 = \sqrt{1 + (A_4 \sigma_{\xi_1})^2}$ and $\bar{\nu}_3 = \frac{A_3}{k_7}$.

Thus, we can rewrite equation (4.28) after replacing its components with the corresponding ones in (4.29), (4.30) and (4.31), and the final form of this equation is:

$$
E[Y_1 \operatorname{expit}\{\theta_0^* + \theta_1^* Y_1\}] \approx \beta_1 \, \Phi \left[\frac{A_3}{\sqrt{1 + (A_4 \sigma_{\xi_1})^2}} \right] + \beta_2 \frac{A_6 \, \sigma_x^2}{k_6} \phi(\bar{\nu}_2) + \frac{A_4 \, \sigma_{\xi_1}^2}{k_7} \phi(\bar{\nu}_3). \tag{4.32}
$$

Again, we can rewrite equation (4.10) to be:

$$
\begin{aligned}
E[Y_1 R] &\quad -E[Y_1 \operatorname{expit} \{\theta_0^* + \theta_1^* Y_1\}] \\
&\approx \quad \beta_1 \Phi \left[\frac{c \, \theta_0}{\sqrt{1 + (c \, \theta_1 \, \sigma_U)^2}} \right] + \frac{c \, \sigma_U^2 \, \theta_1}{\sqrt{1 + (c \, \theta_1 \, \sigma_U)^2}} \phi(\bar{\nu}_1) \\
&\quad - \left(\beta_1 \, \Phi \left[\frac{A_3}{\sqrt{1 + (A_4 \sigma_{\xi_1})^2}} \right] + \beta_2 \frac{A_6 \, \sigma_x^2}{k_6} \phi(\bar{\nu}_2) + \frac{A_4 \, \sigma_{\xi_1}^2}{k_7} \phi(\bar{\nu}_3) \right) = 0. \tag{4.33}
\end{aligned}
$$

We need to solve equations (4.21) and (4.33) simultaneously using numerical methods such as Newton-Raphson, to obtain the least false values of θ_0 and θ_1. This will be illustrated later. In the following we will use these values to calculate the least false values β_3^* and β_4^*.

4.4.3   Calculating the least false values β_3^* and β_4^*

Assuming we have the least false values θ_0^* and θ_1^*, we can treat them as constants. Recall equation (4.6), in order to find the least false values β_3^* and β_4^* we need to calculate the following: $E[R/\pi]$, $E[Rx/\pi]$, $E[Rx^2/\pi]$, $E[RY_2/\pi]$ and $E[RY_2 x/\pi]$. Methods are similar to those used previously and so the calculations and final results are given in Appendix A.4.

4.5   Performance under MNAR

Under MNAR as the true model and MAR as the assumed model, we will estimate the least false values θ_0^* and θ_1^* from equations (4.9) and (4.10), using an approach similar to that used in Chapter 3, and the previous section.

4.5.1   First: Working on equation (4.9)

In the following we will work on equation (4.9) to find the expected values of its components. We start with the first part of equation (4.9), $E[R]$. Since under MNAR R depends on x, U and ϵ_2, then we can find the expectation by using the following approach. First, we find the conditional expectation when x is treated as a constant. We have

$$
\begin{aligned}
E[R|x] &= E[\text{ expit }\{\theta_0 + \theta_2 Y_2\}] \\
&\approx E[\Phi[c(\theta_0 + \theta_2 Y_2)]] \\
&= E[\Phi[c(\theta_0 + \theta_2(\beta_3 + \beta_4 x + U + \epsilon_2))]].
\end{aligned}
\tag{4.34}
$$

$$
\text{Let } U + \epsilon_2 = \xi_2, \text{therefore } \xi_2 \sim N(0, \sigma_{\xi_2}^2), \text{ where } \sigma_{\xi_2}^2 = \sigma_U^2 + \sigma_{\epsilon_2}^2.
\tag{4.35}
$$

So

$$
\begin{aligned}
E[R|x] &\approx E[\Phi[c(\theta_0 + \theta_2(\beta_3 + \beta_4 x)) + c\,\theta_2\xi_2]] \\
&= E[\Phi[w_1 + w_2\xi_2]] \\
&= \int \Phi[w_1 + w_2\xi_2]\phi(\xi_2; 0, \sigma_{\xi_2}^2)\, \mathrm{d}\xi_2 \\
&= \Phi\left[\frac{w_1}{\sqrt{1 + (w_2\sigma_{\xi_2})^2}}\right] \\
&= \Phi\left[\frac{c(\theta_0 + \theta_2\beta_3)}{\sqrt{1 + (c\theta_2\sigma_{\xi_2})^2}} + \frac{c\theta_2\beta_4 x}{\sqrt{1 + (c\theta_2\sigma_{\xi_2})^2}}\right] \\
&= \Phi[B_1 + B_2 x], \text{ say,}
\end{aligned}
$$

$$\tag{4.36}$$
$$\tag{4.37}$$
$$\tag{4.38}$$
$$\tag{4.39}$$

where $w_1 = c(\theta_0 + \theta_2(\beta_3 + \beta_4 x))$ and $w_2 = c\,\theta_2$ are now used as working notations.

Since $E[R] = E\left[E[R|x]\right]$, then

$$
\begin{aligned}
E[R] &\approx E\left[\Phi[B_1 + B_2 x]\right] \\
&= \Phi\left[\frac{B_1}{\sqrt{1 + (B_2\sigma_x^2)^2}}\right].
\end{aligned}
\tag{4.40}
$$

The second part of equation (4.9) is:

$$
E[\text{ expit }\{\theta_0^* + \theta_1^* Y_1\}] = E[R^*].
\tag{4.41}
$$

To calculate these expectations, as before we need to do this in two steps. First assume x is treated as a constant, leading to

$$
\begin{aligned}
E[R^*|x] &= E[\ \text{expit}\ \{\theta_0^* + \theta_1^*(\beta_1 + \beta_2 x) + \theta_1^*(U + \epsilon_1)\}] \\
&\approx E[\Phi[c\,(\theta_0^* + \theta_1^*(\beta_1 + \beta_2 x)) + c\,\theta_1^*\xi_1] \\
&= E[\Phi(w_3 + w_4\xi_1)],
\end{aligned}
$$

where $w_3 = c\,(\theta_0^* + \theta_1^*(\beta_1 + \beta_2 x))$ and $w_4 = c\,\theta_1^*$, and we recall $U + \epsilon_1 = \xi_1$ as defined in equation (4.18). Then

$$
\begin{aligned}
E[R^*|x] &\approx \int \Phi[w_3 + w_4\xi_1]\,\phi(\xi_1; 0, \sigma_{\xi_1}^2)\mathrm{d}\xi_1 \\
&= \Phi\left[\frac{w_3}{\sqrt{1 + (w_4\sigma_{\xi_1})^2}}\right] \\
&= \Phi\left[\frac{c\,(\theta_0^* + \theta_1^*(\beta_1 + \beta_2 x))}{\sqrt{1 + (c\,\theta_1^*\sigma_{\xi_1})^2}}\right] \\
&= \Phi[B_3 + B_4\,x],
\end{aligned}
\tag{4.42}
$$

where $B_3 = \frac{c\,(\theta_0^* + \theta_1^*\beta_1)}{\sqrt{1 + (c\,\theta_1^*\sigma_{\xi_1})^2}}$ and $B_4 = \frac{c\,\theta_1^*\beta_2}{\sqrt{1 + (c\,\theta_1^*\sigma_{\xi_1})^2}}$.

Then we integrate over x to find

$$
\begin{aligned}
E[R^*] = E[E[R^*|x]] &\approx E\left[\Phi[B_3 + B_4\,x]\right] \\
&= \int \Phi[B_3 + B_4\,x]\,\phi(x; 0, \sigma_x^2)\mathrm{d}x \\
&= \Phi\left[\frac{B_3}{\sqrt{1 + (B_4\sigma_x)^2}}\right].
\end{aligned}
\tag{4.43}
$$

Finally, we can replace the values of equations (4.40) and (4.43) with the corresponding values in equation (4.9) to give the first estimating equation

$$
E[R] - E[\text{expit}\ \{\theta_0^* + \theta_1^* Y_1\}] = 0,
$$

or equivalently,

$$
E[R] - E[R^*] \approx \Phi\left[\frac{B_1}{\sqrt{1 + (B_2\sigma_x^2)^2}}\right] - \Phi\left[\frac{B_3}{\sqrt{1 + (B_4\sigma_x)^2}}\right] = 0.
\tag{4.44}
$$

4.5.2   Second: Working on equation (4.10)

In the following we will work on equation (4.10) starting with its first part

$$
\begin{aligned}
E[Y_1 R] &= E[Y_1\ \text{expit}\ \{\theta_0 + \theta_2 Y_2\}] \\
&= E[(\beta_1 + \beta_2\,x + U + \epsilon_1)\ \text{expit}\ \{\theta_0 + \theta_2(\beta_3 + \beta_4 + U + \epsilon_2)\}].
\end{aligned}
$$

We split this into four parts:

$$
\begin{aligned}
E[Y_1 R] &= E[\beta_1 \, \text{expit} \, \{\theta_0 + \theta_2(\beta_3 + \beta_4 \, x + U + \epsilon_2)\}] \\
&+ E[\beta_2 \, x \, \text{expit} \, \{\theta_0 + \theta_2(\beta_3 + \beta_4 x + U + \epsilon_2)\}] \\
&+ E[U \, \text{expit} \, \{\theta_0 + \theta_2(\beta_3 + \beta_4 x + U + \epsilon_2)\}] \\
&+ E[\epsilon_1 \, \text{expit} \, \{\theta_0 + \theta_2(\beta_3 + \beta_4 x + U + \epsilon_2)\}] \\
&= F_1 + F_2 + F_3 + F_4.
\end{aligned} \tag{4.45}
$$

In the following we will find each part separately. Firstly

$$
\begin{aligned}
F_1 &= E[\beta_1 \, \text{expit} \, \{\theta_0 + \theta_2(\beta_3 + \beta_4 \, x + U + \epsilon_2)\}] \\
&= E[\beta_1 \, \text{expit} \, \{\theta_0 + \theta_2(\beta_3 + \beta_4 \, x + \xi_2)\}] \\
&\approx \beta_1 E\left[\Phi[c\,(\theta_0 + \theta_2(\beta_3 + \beta_4 \, x + \xi_2))]\right] \\
&= \beta_1 E\left[\Phi[c\,(\theta_0 + \theta_2(\beta_3 + \beta_4 \, x) + c\,\theta_2\xi_2]\right] \\
&= \beta_1 E\left[\Phi[w_5 + w_6 \, x + w_7 \, \xi_2]\right] \\
&= \beta_1 \int \int \Phi[w_5 + w_6 \, x + w_7 \, \xi_2] \, \phi(\xi_2; 0, \sigma_{\xi_2}^2) \, \phi(x; 0, \sigma_x^2) \mathrm{d}\xi_2 \mathrm{d}x \\
&= \beta_1 \int \left(\int \Phi[w_5 + w_6 \, x + w_7 \, \xi_2] \, \phi(\xi_2; 0, \sigma_{\xi_2}^2) \mathrm{d}\xi_2 \right) \phi(x; 0, \sigma_x^2) \, \mathrm{d}x \\
&= \beta_1 \int \left(\Phi\left[\frac{w_5 + w_6 \, x}{\sqrt{1 + (w_7 \, \sigma_{\xi_2})^2}} \right] \right) \phi(x; 0, \sigma_x^2) \, \mathrm{d}x \\
&= \beta_1 \int \Phi[B_5 + B_6 \, x] \, \phi(x; 0, \sigma_x^2) \, \mathrm{d}x \\
&= \beta_1 \Phi\left[\frac{B_5}{\sqrt{1 + (B_6 \, \sigma_x)^2}} \right],
\end{aligned} \tag{4.46}
$$

where $B_5 = \frac{w_5}{\sqrt{1+(w_7 \, \sigma_{\xi_2})^2}}$, $B_6 = \frac{w_6}{\sqrt{1+(w_7 \, \sigma_{\xi_2})^2}}$, $w_5 = c\,(\theta_0 + \theta_2 \, \beta_3)$, $w_6 = c\,(\theta_2 \, \beta_4)$ and $w_7 = c\,\theta_2$.

Secondly,

$$
\begin{aligned}
F_2 &= E[\beta_2\, x \,\text{expit}\,\{\theta_0 + \theta_2(\beta_3 + \beta_4 x + U + \epsilon_2)\}] && (4.47)\\
&\approx E[\beta_2\, x \Phi[c\,(\theta_0 + \theta_2(\beta_3 + \beta_4\, x + \xi_2))]]] \\
&= \beta_2 E\,[x\,\Phi[w_5 + w_6\, x + w_7\, \xi_2]] \\
&= \beta_2 \int \int x\,\Phi[w_5 + w_6\, x + w_7\, \xi_2]\,\phi(\xi_2; 0, \sigma_{\xi_2}^2)\,\phi(x; 0, \sigma_x^2)\mathrm{d}\xi_2\mathrm{d}x \\
&= \beta_2 \int x \left(\int \Phi[w_5 + w_6\, x + w_7\, \xi_2]\,\phi(\xi_2; 0, \sigma_{\xi_2}^2)\mathrm{d}\xi_2\right)\phi(x; 0, \sigma_x^2)\,\mathrm{d}x \\
&= \beta_2 \int x \left(\Phi\left[\frac{w_5 + w_6\, x}{\sqrt{1 + (w_7\, \sigma_{\xi_2})^2}}\right]\right)\phi(x; 0, \sigma_x^2)\,\mathrm{d}x \\
&= \beta_2 \int x\,\Phi[B_5 + B_6\, x]\,\phi(x; 0, \sigma_x^2)\,\mathrm{d}x. && (4.48)
\end{aligned}
$$

Then we can apply formula (A.7), found in Appendix A.1, on equation (4.48), and this gives:

$$
F_2 \approx \beta_2\frac{\sigma_x^2 B_6}{\sqrt{1 + (B_6\sigma_x)^2}}\phi(\frac{B_5}{\sqrt{1 + (B_6\sigma_x)^2}}). \tag{4.49}
$$

And similarly, F_3 is found as

$$
\begin{aligned}
F_3 &= E[U \,\text{expit}\,\{\theta_0 + \theta_2(\beta_3 + \beta_4 x + U + \epsilon_2)\}] \\
&\approx E[U\Phi[c(\theta_0 + \theta_2(\beta_3 + U) + c\theta_2(\beta_4\, x + \epsilon_2)]].
\end{aligned}
$$

Let $\beta_4\, x + \epsilon_2 = \xi_3$, thus $\xi_3 \sim N(0, \sigma_{\xi_3}^2)$ where $\sigma_{\xi_3}^2 = (\beta_4)^2\sigma_x^2 + \sigma_{\epsilon_2}^2$. So $\qquad\qquad$ (4.50)

$$
\begin{aligned}
F_3 &\approx E[U\Phi[c(\theta_0 + \theta_2(\beta_3 + U) + c\theta_2\, \xi_3]] \\
&= E[U\Phi[w_5 + w_7\, \xi_3 + w_7\, U]] \\
&= \int \int U\,\Phi[w_5 + w_7\, \xi_3 + w_7\, U]\,\phi(\xi_3; 0, \sigma_{\xi_3}^2)\,\phi(U; 0, \sigma_U^2)\mathrm{d}\xi_3\mathrm{d}U \\
&= \int U \left(\int \Phi[w_5 + w_7\, \xi_3 + w_7\, U]\,\phi(\xi_3; 0, \sigma_{\xi_3}^2)\,\mathrm{d}\xi_3\right)\phi(U; 0, \sigma_U^2)\,\mathrm{d}U \\
&= \int U \left(\int \Phi[w_8 + w_7\, \xi_3]\,\phi(\xi_3; 0, \sigma_{\xi_3}^2)\,\mathrm{d}\xi_3\right)\phi(U; 0, \sigma_U^2)\,\mathrm{d}U \\
&= \int U \left(\Phi\left[\frac{w_8}{\sqrt{1 + (w_7\, \sigma_{\xi_3})^2}}\right]\right)\phi(U; 0, \sigma_U^2)\,\mathrm{d}U \\
&= \int U\,\Phi[B_7 + B_8\, U]\,\phi(U; 0, \sigma_U^2)\,\mathrm{d}U \\
&= \frac{\sigma_U^2 B_8}{\sqrt{1 + (B_8\sigma_U)^2}}\phi(\frac{B_7}{\sqrt{1 + (B_8\sigma_U)^2}}), && (4.51)
\end{aligned}
$$

where $B_7 = \frac{w_5}{\sqrt{1+(w_7\,\sigma_{\xi_3})^2}}$, $B_8 = \frac{w_7}{\sqrt{1+(w_7\,\sigma_{\xi_3})^2}}$ and $w_8 = w_5 + w_7\,U$.

Finally, for F_4 we note that the random variable ϵ_1 is independent of x and U and as it has a mean of zero.

$$F_4 = E[\epsilon_1 \ \text{expit}\ \{\theta_0 + \theta_2(\beta_3 + \beta_4 x + U + \epsilon_2)\}] = 0. \tag{4.52}$$

Using equations from (4.46) to (4.52) to calculate $E[Y_1 R]$ in equation (4.45), we find

$$E[Y_1 R] \approx \beta_1 \Phi\left[\frac{B_5}{\sqrt{1+(B_6\,\sigma_x)^2}}\right] + \beta_2 \frac{\sigma_x^2 B_6}{\sqrt{1+(B_6\sigma_x)^2}}\phi(\bar{\nu}_4) + \frac{\sigma_U^2 B_8}{\sqrt{1+(B_8\sigma_U)^2}}\phi(\bar{\nu}_5), \tag{4.53}$$

where $\bar{\nu}_4 = \frac{B_5}{\sqrt{1+(B_6\sigma_x)^2}}$ and $\bar{\nu}_5 = \frac{B_7}{\sqrt{1+(B_8\sigma_U)^2}}$.

We will now work on the second part of equation (4.10). First

$$E[Y_1 \text{expit}\{\theta_0^* + \theta_1^* Y_1\}] \quad = \quad E[(\beta_1 + \beta_2 + U + \epsilon_1)\,\text{expit}\{\theta_0^* + \theta_1^*(\beta_1 + \beta_2\,x + U + \epsilon_1)\}].$$

As $U + \epsilon_1 = \xi_1$,

$$
\begin{aligned}
E[Y_1 \text{expit}\{\theta_0^* + \theta_1^* Y_1\}] \quad = \quad & E[\beta_1 \text{expit}\{\theta_0^* + \theta_1^*(\beta_1 + \beta_2\,x + \xi_1)\}] \\
+ \quad & E[\beta_2\,x\,\text{expit}\{\theta_0^* + \theta_1^*(\beta_1 + \beta_2\,x + \xi_1)\}] \\
+ \quad & E[\xi_1\text{expit}\{\theta_0^* + \theta_1^*(\beta_1 + \beta_2\,x + \xi_1)\}] \\
= \quad & F_5 + F_6 + F_7. \tag{4.54}
\end{aligned}
$$

The first of these is

$$
\begin{aligned}
F_5 \quad = \quad & E[\beta_1 \text{expit}\{\theta_0^* + \theta_1^*(\beta_1 + \beta_2\,x + \xi_1)\}] \\
\approx \quad & \beta_1 E[\Phi[c\,(\theta_0^* + \theta_1^*(\beta_1 + \beta_2\,x + \xi_1))]] \\
= \quad & \beta_1 E[\Phi[c\,(\theta_0^* + \theta_1^*(\beta_1 + \beta_2\,x)) + c\,\theta_1^*\xi_1]] \\
= \quad & \beta_1 E[\Phi[w_9 + w_{10}\xi_1]] \\
= \quad & \beta_1 \int \int \Phi[w_9 + w_{10}\,x + w_{11}\xi_1]\phi(x;0,\sigma_x^2)\phi(\xi_1;0,\sigma_{\xi_1}^2)\,\mathrm{d}x\,\mathrm{d}\xi_1 \\
= \quad & \beta_1 \int \left(\int \Phi[w_9 + w_{10}\,x + w_{11}\xi_1]\phi(x;0,\sigma_x^2)\,\mathrm{d}x\right)\phi(\xi_1;0,\sigma_{\xi_1}^2)\mathrm{d}\xi_1 \\
= \quad & \beta_1 \int \left(\int (\Phi[w_{10}\,x + w_{12}]\phi(x;0,\sigma_x^2)\,\mathrm{d}x\right)\phi(\xi_1;0,\sigma_{\xi_1}^2)\mathrm{d}\xi_1.
\end{aligned}
$$

We will do the inner integration first using the ESN formula (A.2) in Appendix A.1, thus

$$
\begin{aligned}
F_5 &\approx \beta_1 \int \left(\varPhi \left[\frac{w_{12}}{\sqrt{1+(w_{10}\sigma_x)^2}} \right] \right) \phi(\xi_1; 0, \sigma_{\xi_1}^2) \mathrm{d}\xi_1 \\
&= \beta_1 \int \left(\varPhi \left[\frac{w_9 + w_{11}\xi_1}{\sqrt{1+(w_{10}\sigma_x)^2}} \right] \right) \phi(\xi_1; 0, \sigma_{\xi_1}^2) \mathrm{d}\xi_1 \\
&= \beta_1 \int \varPhi \left[B_9 + B_{10}\xi_1 \right] \phi(\xi_1; 0, \sigma_{\xi_1}^2) \mathrm{d}\xi_1.
\end{aligned}
$$

Again, this integral can be calculated using the same formula; the final result is:

$$
F_5 \approx \beta_1 \varPhi \left[\frac{B_9}{\sqrt{1+(B_{10}\sigma_{\xi_1})^2}} \right], \tag{4.55}
$$

where $B_9 = \frac{w_9}{\sqrt{1+(w_{10}\sigma_x)^2}}$, $B_{10} = \frac{w_{11}}{\sqrt{1+(w_{10}\sigma_x)^2}}$, $w_9 = c\,(\theta_0^* + \theta_1^*\beta_1)$, $w_{10} = c\,\theta_1^*\beta_2$, and $w_{11} = c\,\theta_1^*$.

Turning to F_6 and F_7, the final results are

$$
F_6 \approx \beta_2 \frac{B_{12}\,\sigma_x^2}{w_{14}} \phi(\bar{\nu}_6) \tag{4.56}
$$

$$
F_7 \approx \frac{B_{10}\,\sigma_{\xi_1}^2}{w_{15}} \phi(\bar{\nu}_7), \tag{4.57}
$$

where $B_{11} = \frac{w_9}{\sqrt{1+(w_{11}\sigma_{\xi_1})^2}}$, $B_{12} = \frac{w_{10}}{\sqrt{1+(w_{11}\sigma_{\xi_1})^2}}$, $w_{13} = w_9 + w_{10}\,x$, $w_{14} = \sqrt{1+(B_{12}\sigma_x)^2}$, $w_{15} = \sqrt{1+(B_{10}\sigma_{\xi_1})^2}$, $\nu_6 = \frac{B_{11}}{w_{14}}$ and $\bar{\nu}_7 = \frac{B_9}{w_{15}}$.

Thus, we can rewrite equation (4.54) after replacing its components by the corresponding ones in (4.55), (4.56) and (4.57) and the final form of this equation is:

$$
E[Y_1 \exp\mathrm{it}\{\theta_0^* + \theta_1^* Y_1\}] \approx \beta_1 \varPhi \left[\frac{B_9}{\sqrt{1+(B_{10}\sigma_{\xi_1})^2}} \right] + \beta_2 \frac{B_{12}\,\sigma_x^2}{w_{14}} \phi(\bar{\nu}_6) + \frac{B_{10}\,\sigma_{\xi_1}^2}{w_{15}} \phi(\bar{\nu}_7). \tag{4.58}
$$

We now turn to equation (4.10) under MNAR using equation (4.53) and equation (4.58), leading to the second estimating equation

$$
\begin{aligned}
E[Y_1 R] \;&-\; E[Y_1 \exp\mathrm{it}\,\{\theta_0^* + \theta_1^* Y_1\}] \\
\approx\quad &\beta_1 \varPhi \left[\frac{B_5}{\sqrt{1+(B_6\,\sigma_x)^2}} \right] + \beta_2 \frac{\sigma_x^2 B_6}{\sqrt{1+(B_6\sigma_x)^2}} \phi(\bar{\nu}_4) + \frac{\sigma_U^2 B_8}{\sqrt{1+(B_8\sigma_U)^2}} \phi(\bar{\nu}_5) \\
-\quad &\left(\beta_1 \varPhi \left[\frac{B_9}{\sqrt{1+(B_{10}\sigma_{\xi_1})^2}} \right] + \beta_2 \frac{B_{12}\,\sigma_x^2}{w_{14}} \phi(\bar{\nu}_6) + \frac{B_{10}\,\sigma_{\xi_1}^2}{w_{15}} \phi(\bar{\nu}_7) \right) = 0. \tag{4.59}
\end{aligned}
$$

Then we can solve equations (4.44) and (4.59) simultaneously using numerical methods such as Newton-Raphson, as mentioned in Section 4.4.2, which gives us the limiting values θ_0^* and θ_1^*.

4.5.3 Calculating the limiting values β_3^* and β_4^*

Assuming we have got the least false values θ_0^* and θ_1^* from the previous section, hence we can treat them as constants. Recall equation (4.6), in order to find the least false values β_3^* and β_4^* we need to calculate the following: $E[R/\pi]$, $E[Rx/\pi]$, $E[Rx^2/\pi]$, $E[RY_2/\pi]$ and $E[RY_2x/\pi]$. Unfortunately, there is no closed approximation for these terms, as shown in Appendix A.5. In the numerical work to follow, we obtain the expectations using Monte Carlo methods.

4.6 Simulations and Numerical Investigation

We will do this numerical investigation in two parts as in the previous chapter. In the first part we want to check that the theory and our least false values are correct by comparing the expected values with the simulated ones. In the second part we explore in detail how the misspecification affects the β estimates, using the theoretical results. Now that we have working formulas there is no need for simulations here except under MNAR.

As before we generated a scalar $N(0,1)$ variable x, then we generated the longitudinal means $\mu_1 = \beta_1 + \beta_2 x$, $\mu_2 = \beta_3 + \beta_4 x$. This was followed by (Y_1, Y_2) from a bivariate normal distribution with mean (μ_1, μ_2). In all of the following simulations, unless it is stated otherwise, we take the same parameters as used in the previous chapter. Our choice of parameters will keep the dropout always about 50%.

Firstly, we want to know at what n can we be confident that the theory is reasonable. We compare the theoretical results with the simulated means as n increases; the results are shown in Table 4.1 and Table 4.2. We will see that the simulated values are close to the corresponding expected values, so indicating that the limiting values found by the theory are correct for sample size not less than about 100000. Based on this result for each set of parameters we studied a sample size of 1000000. In addition, the 95% reference intervals for the least false values under SP, MAR and MNAR are shown in Table 4.3 to Table 4.4.

Secondly, as in the previous chapter, for big samples, we will try to find several combinations of $(\theta_0, \theta_1, \theta_2)$ under fixed β and for a given variances $\sigma_x^2 = 1$, $\sigma_U = \sigma_{\epsilon_1} = \sigma_{\epsilon_2} = \sqrt{0.5}$ to keep the total amount of the dropout fixed and at about 50%. These combinations of θs are shown in Table 4.5 and Table 4.6. In addition, we calculate the limiting values for different θ combinations. Moreover, we will suppose that we have a given θ (and dropout), and we will study the effect of changing the variance parameters. The results are shown in Table 4.7.

Finally we study the effect of changing θ on β_3^* and β_4^* by producing contour plots shown

in Figures 4.1 to 4.3.

4.6.1 The limiting values compared to the simulated values for different sample size

Here we attempt to show what happens as sample size increases, i.e. at what n can we be confident that the theory is reasonable. One can note that as expected as n increases the simulated and theoretical values become close to each other. Note that $\hat{\theta}_0$ and $\hat{\theta}_1$ are the estimated values produced by the simulations when we fit MAR to data that are MNAR or SP.

Results at $n{=}100000$ show that the simulated values all match the theoretical approximations very well, which supports the use of the theory. Comparing the result here with the corresponding result in Chapter 3, we note that here we need a larger sample size to achieve n that gives us an acceptable result. This may be related to the extra variability of IPW observed in Chapter 2.

Table 4.1: The limiting values under different sample sizes

n		SP	MAR	MNAR
1000	θ_0^*	0.14	1.00	0.48
	θ_1^*	0.07	0.50	0.24
	$\hat{\theta}_0$	0.05	1.03	0.47
	$\hat{\theta}_1$	0.06	0.54	0.22
	β_3^*	-0.85	-1.00	-0.82
	β_4^*	-1.00	-1.01	-0.96
	$\hat{\beta}_3$	-0.76	-1.03	-0.75
	$\hat{\beta}_4$	-0.99	-1.04	-0.94
10000	θ_0^*	0.14	1.00	0.48
	θ_1^*	0.07	0.50	0.24
	$\hat{\theta}_0$	0.15	0.97	0.49
	$\hat{\theta}_1$	0.07	0.50	0.25
	β_3^*	-0.85	-1.00	-0.82
	β_4^*	-1.00	-1.01	-0.96
	$\hat{\beta}_3$	-0.85	-1.02	-0.80
	$\hat{\beta}_4$	-0.99	-1.01	-0.95

Table 4.2: The limiting values under different sample sizes

n		SP	MAR	MNAR
50000	θ_0^*	0.14	1.00	0.48
	θ_1^*	0.07	0.50	0.24
	$\hat{\theta}_0$	0.14	1.02	0.48
	$\hat{\theta}_1$	0.07	0.50	0.24
	β_3^*	-0.85	-1.00	-0.82
	β_4^*	-1.00	-1.01	-0.96
	$\hat{\beta}_3$	-0.85	-0.99	-0.81
	$\hat{\beta}_4$	-1.01	-0.98	-0.96
100000	θ_0^*	0.14	1.00	0.48
	θ_1^*	0.07	0.50	0.24
	$\hat{\theta}_0$	0.14	1.00	0.48
	$\hat{\theta}_1$	0.07	0.50	0.24
	β_3^*	-0.85	-1.00	-0.82
	β_4^*	-1.00	-1.01	-0.96
	$\hat{\beta}_3$	-0.85	-1.00	-0.82
	$\hat{\beta}_4$	-1.00	-1.00	-0.96
500000	θ_0^*	0.14	1.00	0.48
	θ_1^*	0.07	0.50	0.24
	$\hat{\theta}_0$	0.14	1.00	0.48
	$\hat{\theta}_1$	0.07	0.50	0.24
	β_3^*	-0.85	-1.00	-0.82
	β_4^*	-1.00	-1.00	-0.96
	$\hat{\beta}_3$	-0.85	-1.00	-0.82
	$\hat{\beta}_4$	-1.00	-1.00	-0.96

4.6.2   The 95% reference interval for the limiting values

We simulate 100 samples of size 100000, then find the $\hat{\beta}$ for each sample then calculate the reference interval instead of standard confidence interval as mentioned in Chapter 3 Section 3.4.2, so we will say: CI=$\bar{\hat{\beta}} \pm 2\,SD(\beta)$. We did this for the SP, MAR and MNAR dropout models, each with just one set of dropout parameters and all with the missingness percentage at about 50% dropout. Results show that the reference ranges all include the theoretical values.

Table 4.3: The 95% reference interval for β_3^* using 100 samples of size 100000.

	TV	SV	Lower Bound	Upper Bound
MAR	-0.9920	-1.0009	-1.0114	-0.9903
MNAR	-0.8274	-0.8207	-0.8305	-0.8110
SP	-0.8285	-0.8472	-0.8533	-0.8411

Table 4.4: The 95% reference interval for β_4^* using 100 samples of size 100000.

	TV	SV	Lower Bound	Upper Bound
MAR	-0.9928	-1.0015	-1.0162	-0.9868
MNAR	-0.9638	-0.9591	-0.9697	-0.9486
SP	-1.0055	-1.0015	-1.0121	-0.9909

4.6.3 Effect on dropout parameters of fitting a MAR model when the true model is MNAR or SP

In Table 4.5, we show θ_0^* and θ_1^* by solving (4.21) and (4.33) for the given values of θ_0^{sp} and θ_1^{sp} with the same parameters used to generate the logistic model as mentioned before. These values of θ all give a dropout rate of around 50%. The columns for β_3^* and β_4^* give the least false value for β_3 and β_4, respectively.

In Table 4.6, we show θ_0^* and θ_1^* by solving (4.44) and (4.59) for the given values of θ_0^{MN} and θ_2^{MN} and $\theta_1^{MN} = 0$ with the same parameters used to generate the logistic model as mentioned before. These values of θ all give a dropout rate of around 50%. The columns for β_3^* and β_4^* give the least false value for β_3 and β_4, respectively.

Recall that the true values are $\beta_3^G = \beta_4^G = -1$.

Table 4.5: Effect of fitting a MAR model when the true model is SP

θ_0^{sp}	θ_1^{sp}	θ_0^*	θ_1^*	β_3^*	β_4^*
0	-1.414	-0.247	-0.121	-1.275	-1.000
0	-0.707	- 0.132	- 0.066	-1.155	-1.002
0	0.707	0.132	0.066	-0.843	-0.999
0	1.414	0.247	0.121	-0.743	-1.000

Table 4.6: Effect of fitting a MAR model when the true model is MNAR

θ_0^{MN}	θ_2^{MN}	θ_0^*	θ_1^*	β_3^*	β_4^*
-1	-1	-0.871	-0.439	-1.330	-0.869
-0.5	-0.5	-0.507	-0.241	-1.180	-0.960
0.5	0.5	0.479	0.239	-0.819	-0.963
1	1	0.891	0.452	-0.672	-0.881

Results from Table 4.5 and Table 4.6 are what we most care about. Under SP (Table 4.5), the least false values of β_4^* are very close to the true quantity. Hence we only investigate the effect of θ on β_3. The bias in β_3 is larger as the absolute value of θ_1^{sp} get bigger. For negative θ_1^{sp}, we note that β_3 tends to be smaller than its true value. The chance of dropout goes up with Y_2 in this case, so we lose the large Y_2 values and the mean of the observed data at time 2 is too low. So the intercept term (β_3) is too low. A correct inverse probability model would compensate for this, but the misspecifed one does not do enough.

The opposite happens for positive θ_1^{sp}.

Turning to the next table, Table 4.6 under MNAR, β_4 is biased towards zero as previously seen (Appendix A.6). But β_3 is biased either down or up depending upon whether dropout is associated with large Y_2 (negative θ_2^{MN}), or small Y_2 (positive θ_2^{MN}) as previously. If dropout is associated with small Y_2, then the mean in the observed data will be too high, and the wrong missingness model does not compensate for this. The opposite is true if dropout is associated with large Y_2.

4.6.4 The effect of the variance of the random effect on the limiting values β_3^* and β_4^*

In the following table we will study the effect of changing the variance of the random effect on the limiting values at different dropout models. We take $\sigma_U^2 = \sigma_{\epsilon_1}^2 = \sigma_{\epsilon_2}^2$. Each time, we will keep the same dropout percentage, which is always about 50%.

Table 4.7: The effect of the variance σ_U^2 on the limiting values β_3^* and β_4^*.

		SP	MAR	MNAR
$\sigma_U=\sqrt{0.5}$	β_3^*	-0.8380	-1.0020	-0.8210
	β_4^*	-0.9950	-0.9860	-0.9560
	$Bias(\beta_3)$	0.1620	0.0020	0.1790
	$Bias(\beta_4)$	0.0050	0.0140	0.0440
$\sigma_U=\sqrt{1}$	β_3^*	-0.7300	-0.9920	-0.6470
	β_4^*	-1.0050	-0.9900	-0.9300
	$Bias(\beta_3)$	0.2700	0.0080	0.3530
	$Bias(\beta_4)$	0.0050	0.0100	0.0700
$\sigma_U=\sqrt{3}$	β_3^*	-0.3990	-1.0310	-0.0930
	β_4^*	-1.0320	-1.0230	-0.8470
	$Bias(\beta_3)$	0.6010	0.0310	0.9070
	$Bias(\beta_4)$	0.0320	0.0230	0.1530
$\sigma_U=\sqrt{6}$	β_3^*	-0.0220	-1.0450	0.6100
	β_4^*	-1.0830	-1.0350	-0.7180
	$Bias(\beta_3)$	0.9780	0.0450	1.6100
	$Bias(\beta_4)$	0.0830	0.0350	0.2820
$\sigma_U=\sqrt{12}$	β_3^*	0.6140	-0.9470	1.7070
	β_4^*	-1.1640	-0.9630	-0.6090
	$Bias(\beta_3)$	1.6140	0.0530	2.7070
	$Bias(\beta_4)$	0.1640	0.0370	0.3910

As for Chapter 3, changing the variance parameters and keeping the same dropout percentage at about 50% has a remarkable effect on the bias. From Table 4.7, it is clear that as σ_u (and consequentially σ_{ϵ_1} and σ_{ϵ_2}) increases, the limiting values β_3^* and β_4^* go further from the true value ($\beta_3^G = -1, \beta_4^G = -1$) and hence the bias increases, which means that the large error variances imply poor results. In short we can conclude that the more variability, the more bias.

4.6.5 Contour plots

Here we want to show how the limiting values β^* change when we change the θs. We produce contour plots of β^* for a grid with $-1.5 < \theta_0^{sp} < 1.5$ and $-1.5 < \theta_1^{sp} < 1.5$. Here we do not present a plot for β_4^* under SP, because the IPW method gives close estimates for β_4 under SP dropout.

We can conclude from the contour plot of β_3^* under SP (Figure 4.1), that in order to minimise the bias in β_3^*, we should choose θ_1 to be around zero, which is equivalent to MCAR. The biggest change is in the vertical direction, for example at $\theta_1 = 0$, then $\beta_3^* = -1$ as expected. For negative θ_1, the dropout is associated with large Y_2, so Y_2 tends to be low. Hence β_3^* is lower than what it should be. The opposite happens for a positive θ_1.

Similarly, we can conclude a typical result from Figure 4.2 of β_3^* under MNAR dropout. For positive θ_2, β_3^* is higher than its correct value, hence Y_2 tends to be high, while β_3^* will be lower for negative θ_2 regardless the value θ_0.

In Figure 4.3 we get positive bias as θ_2 moves away from zero in either direction. This is similar to Figure 3.4 in Chapter 3. Refer to Appendix A.6 for more details.

Figure 4.1: Contour plot of β_3^* under SP.
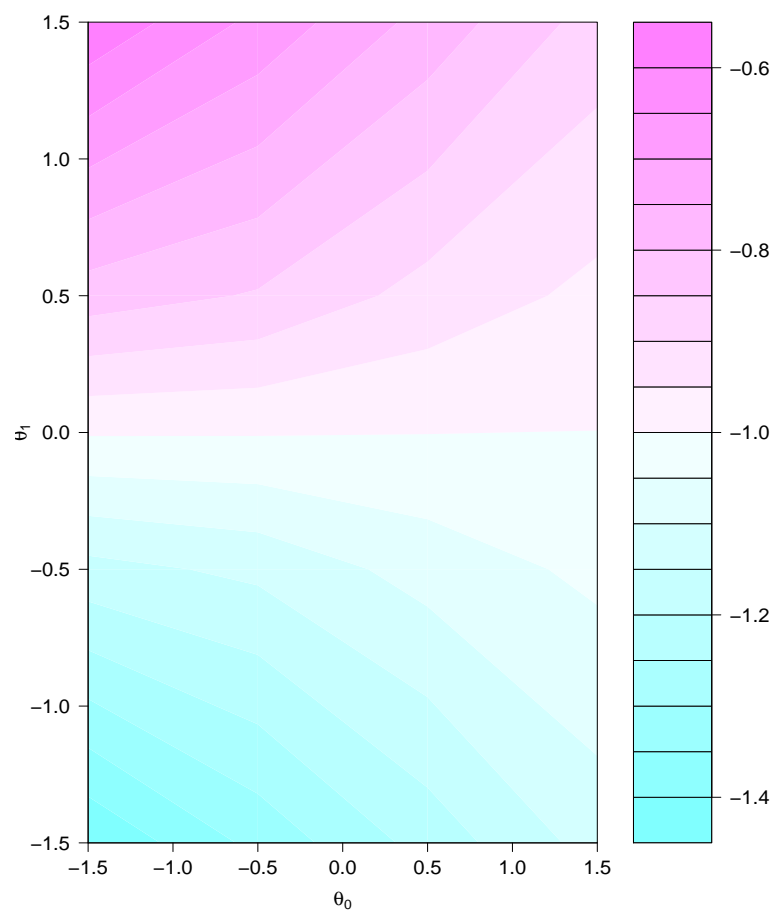
Figure 4.2: Contour plot of β_3^* under MNAR.
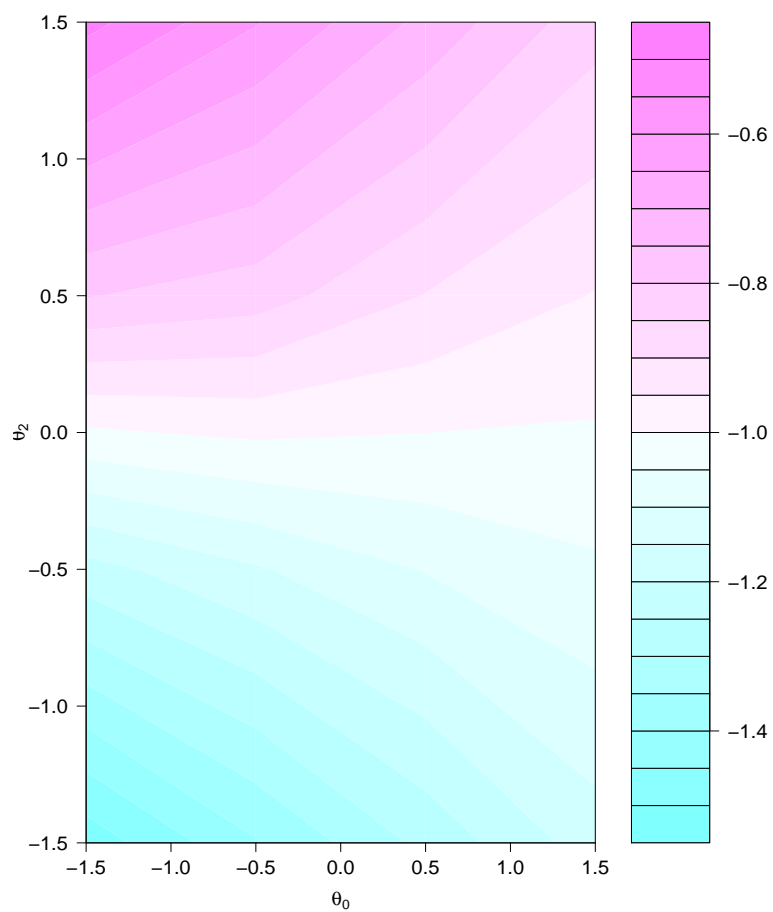
Figure 4.3: Contour plot of β_4^* under MNAR.
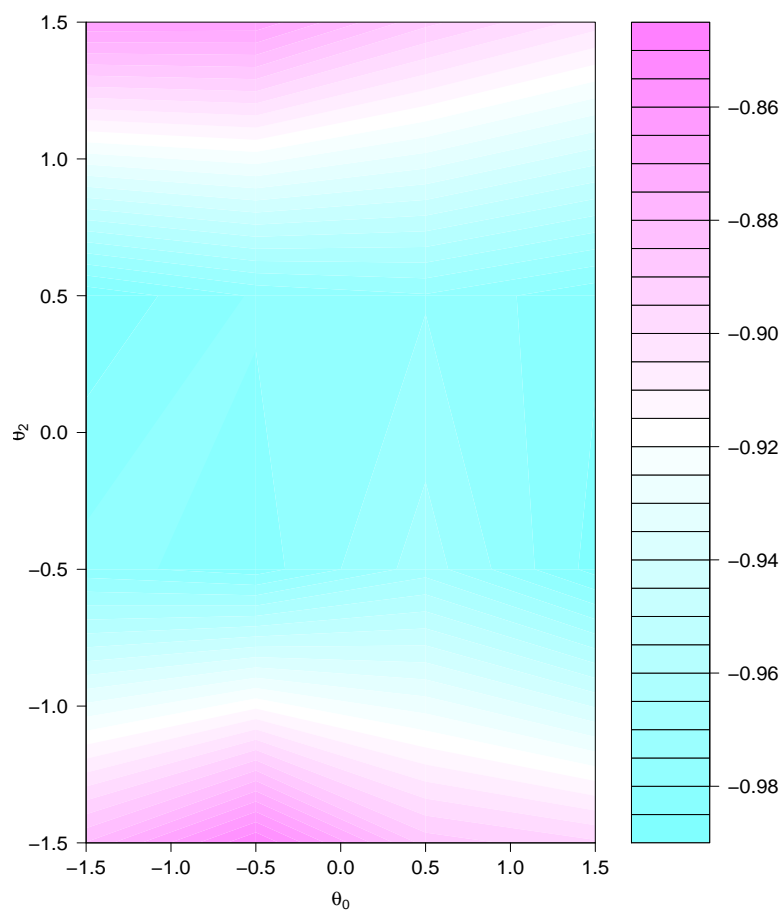
4.7 Discussion and Conclusion

In this chapter we calculated the expected values of terms in equations (4.9) and (4.10), or the least false values of θ_0^* and θ_1^*. Then we calculated the least false values of β_3^* and β_4^* under SP model. We could not find a closed form under MNAR and therefore used numerical results. We compared these expected values under different dropout models to the generating values β_3^G and β_4^G. We calculated the bias in the expected parameters in each dropout model. As we found in Chapter 2 in Tables 2.3 and 2.4, it is clear that the IPW works under missing at random (MAR) since $\beta_3^* = \beta_3^G$ and $\beta_4^* = \beta_4^G$, or, in other words, the bias is zero. In contrast, the IPW method does not seem to work under SP and MNAR as there is bias in both parameters, the simulation results can be found in Chapter 2 in Tables 2.6 and 2.8 for MNAR and SP, respectively. We also conclude that the IPW method gives consistent estimates for β_3 under MAR dropout model and also for β_4 under MAR and SP dropout models. Both of them fail to give a consistent estimate for β_3 or β_4 under MNAR. Results from the contour plots support our conclusions from Table 4.5 and Table 4.6. In the next chapter we will investigate the performance of the CC method and calculate its least false values.

Chapter 5

Performance of CC Method Under Shared Parameter, MAR and MNAR Dropout

5.1 Introduction

In this chapter we investigate the complete case method (CC). Following the methods of the previous chapters, we take the two timepoints situation as a special case and investigate how the CC method performs under shared parameter, MAR and MNAR dropout. As previous, derivation and illustration of theoretical least false values are made under SP and MNAR dropout. To check, we compare the theoretical least false values with simulation results for selected parameter combinations. For a reference see Matthews et al. [2012]. As we saw in Chapter 2, we expect loss of information when the Complete Case method is used and potential bias, except in the case that the data are MCAR.

5.2 Complete Case Method With Two Timepoints

As previous, we assume the generating model for responses is

$$E[Y_{1i}] = \beta_1 + \beta_2 x_i \text{ and } E[Y_{2i}] = \beta_3 + \beta_4 x_i.$$

In complete case analysis only observations with values at both timepoints are used. This means that estimates at time 1 will be affected as well as at time 2. The least squares estimates for time 1 are

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} R_i Y_{1i}(x_i - \overline{Rx}/\overline{R})}{\sum_{i=1}^{n} R_i (x_i - \overline{Rx}/\overline{R})^2} \text{ and } \hat{\beta}_1 = \frac{\overline{RY_1}}{\overline{R}} - \hat{\beta}_2 \frac{\overline{Rx}}{\overline{R}}$$

or

$$\hat{\beta}_2 = \frac{\sum Y_{1i} x_i R_i - \sum Y_{1i} R_i \sum x_i R_i / \sum R_i}{\sum x_i^2 R_i - (\sum x_i R_i)^2 / \sum R_i}.$$

Similarly

$$\hat{\beta}_4 = \frac{\sum_{i=1}^n R_i Y_{2i}(x_i - \overline{Rx}/\overline{R})}{\sum_{i=1}^n R_i(x_i - \overline{Rx}/\overline{R})^2} \text{ and } \hat{\beta}_3 = \frac{\overline{RY_2}}{\overline{R}} - \hat{\beta}_4 \frac{\overline{Rx}}{\overline{R}} \tag{5.1}$$

or

$$\hat{\beta}_4 = \frac{\sum Y_{2i} x_i R_i - \sum Y_{2i} R_i \sum x_i R_i / \sum R_i}{\sum x_i^2 R_i - (\sum x_i R_i)^2 / \sum R_i}. \tag{5.2}$$

The $\overline{Rx}/\overline{R}$ term arises as we use the mean x amongst observed values, which can be written as $\sum_{i=1}^n R_i x_i / \sum_{i=1}^n R_i$. Dividing top and bottom by n allows us to change the sums to means which is helpful for later use of the laws of large numbers (LLN). We will study what happens to the estimates under the three different dropout models, as n increases. We will concentrate on β_3 and β_4, which are the parameters of main interest. Similar methods can be applied for β_1 and β_2, though we need to be careful in defining the dropout model. For example, if the generating model is MAR, dropout (and R) depends on Y_1 but not Y_2. If we choose not to use Y_1 values when Y_2 is missing then this induces a form of MNAR, as missingness depends on the potentially unused Y_1. This would not be sensible - clearly we should use all information in estimating the time 1 parameters. Hence we concentrate on time 2 only.

As we have used several times, as n increases, the LLN implies that $\hat{\beta}_4$ converges to β_4^* given by

$$\hat{\beta}_4 \to \beta_4^* = \frac{E\left\{RY_2(x - E[Rx]/E[R])\right\}}{E\left\{R(x - E[Rx]/E[R])^2\right\}} \tag{5.3}$$

and

$$\hat{\beta}_3 \to \beta_3^* = \frac{E[RY_2]}{E[R]} - \beta_4^* \frac{E[Rx]}{E[R]}. \tag{5.4}$$

To make the calculation easier, we can rewrite β_4^* defined in (5.3) as:

$$\beta_4^* = \frac{E[RY_2 x] - E[RY_2]E[Rx]/E[R]}{E[Rx^2] - (E[Rx])^2/E[R]}. \tag{5.5}$$

Thus, under each dropout model, we need to find $E[R], E[Rx], E[Rx^2], E[RY_2], E[RY_2 x]$, where $E[RY_2] = E[R(\beta_3 + \beta_4 x + U + \epsilon_2)] = \beta_3 E[R] + \beta_4 E[Rx] + E[RU] + E[R\epsilon_2]$, and $E[RY_2 x] = E[Rx(\beta_3 + \beta_4 x + U + \epsilon_2)] = \beta_3 E[Rx] + \beta_4 E[Rx^2] + E[RUx] + E[Rx\epsilon_2]$.

Now let $n \to \infty$. To recall (see Section 3.3.2 in Chapter 3), there are five fundamental variables: U, ϵ_1, ϵ_2, R and x. At times (including this part) it may be more convenient to work with Y_1 and Y_2 instead of U, ϵ_1 and ϵ_2.

5.3   Performance under Shared Parameter Dropout

Under SP, R depends only on U, and as R is independent of x and we know $x \sim N(0, \sigma_x^2)$, then $E[Rx] = E[R]E[x] = 0$ and $E[Rx^2] = E[R]E[x^2] = \sigma_x^2 E[R]$. We found $E[R]$ in equation (4.16) in Chapter 4. Also, $E[RY_2] = \beta_3 E[R] + \beta_4 E[Rx] + E[RU] + E[R\epsilon_2]$.

As $E[Rx] = 0$ and $E[R\epsilon_2] = 0$, we still need to calculate $E[RU]$:

$$
\begin{aligned}
E[RU] &= \int U \text{ expit } \{\theta_0 + \theta_1 U\}\phi(U; 0, \sigma_U^2)\,\mathrm{d}U \\
&\approx \int U\Phi[c(\theta_0 + \theta_1 U)]\phi(U; 0, \sigma_U^2)\,\mathrm{d}U \\
&= \frac{c\,\theta_1\sigma_U^2}{\sqrt{1 + (c\,\theta_1\sigma_U)^2}}\phi\left[\frac{c\,\theta_0}{\sqrt{1 + (c\,\theta_1\sigma_U)^2}}\right].
\end{aligned}
\tag{5.6}
$$

Hence we get $E[RY_2] = \beta_3 E[R] + E[RU]$. Similarly, $E[RY_2 x] = \beta_3 E[Rx] + \beta_4 E[Rx^2] + E[RUx] + E[R\epsilon_2 x]$.

As $E[Rx^2] = \sigma_x^2 E[R]$ and $E[Rx] = E[Rx\epsilon_2] = E[RUx] = 0$, we get $E[RY_2 x] = \beta_4 \sigma_x^2 E[R]$. Hence we have all the terms necessary for β_3^* and β_4^*. The results are illustrated in Section 5.6.

5.4   Performance under MAR Dropout

We found $E[R]$, $E[Rx]$, $E[Rx^2]$ in equations (3.28), (3.29) and (3.34) respectively in Chapter 3, thus we need here to find $E[RY_2]$ and $E[RY_2 x]$.

Consider $E[RY_2] = E[R(\beta_3 + \beta_4 x + U + \epsilon_2)] = \beta_3 E[R] + \beta_4 E[Rx] + E[RU] + E[R\epsilon_2]$, and $E[RY_2 x] = E[Rx(\beta_3 + \beta_4 x + U + \epsilon_2)] = \beta_3 E[Rx] + \beta_4 E[Rx^2] + E[RUx] + E[Rx\epsilon_2]$. Under MAR, R is independent of ϵ_2, thus $E[R\epsilon_2]=0$ and $E[Rx\epsilon_2]=0$. This means we only need to find $E[RU]$ and $E[RUx]$. We start with $E[RU] = E_x[E[RU|x]]$. The conditional expectation is

$$
\begin{aligned}
E[RU|x] &= \int\int U \text{ expit } \{\theta_0 + \theta_1 Y_1\}\phi(U; 0, \sigma_U^2)\phi(\epsilon_1; 0, \sigma_{\epsilon_1}^2)\mathrm{d}U\,\mathrm{d}\epsilon_1 \\
&= \frac{\sigma_U^2 A_4}{\sqrt{1 + A_4^2\sigma_U^2}}\phi\left(\frac{A_3}{\sqrt{1 + A_4^2\sigma_U^2}}\right)
\end{aligned}
$$

where $A_1 = c(\theta_0 + \theta_1\beta_1 + \beta_2 x) + c\theta_1 U$, $A_2 = c\theta_1$, $A_3 = \frac{c(\theta_0 + \theta_1(\beta_1 + \beta_2 x))}{\sqrt{1 + (A_2\sigma_{\epsilon_1})^2}}$, $A_4 = \frac{c\theta_1}{\sqrt{1 + (A_2\sigma_{\epsilon_1})^2}}$.

Then

$$
\begin{aligned}
E[RU] &= E_x[E[RU|x] \\
&= \frac{\sigma_U^2 A_4}{\sqrt{1 + A_4^2 \sigma_U^2}} \int \phi\left(\frac{A_3}{\sqrt{1 + A_4^2 \sigma_U^2}}\right) \phi(x; 0, \sigma_x^2) \mathrm{d}x \\
&= \frac{\sigma_U^2 A_4}{\sqrt{1 + A_4^2 \sigma_U^2}} \frac{e^{-\frac{1}{2} A_7}}{\sqrt{2\pi} \sqrt{A_6^2 + 1}}
\end{aligned}
$$

where $A_5 = \frac{c(\theta_0 + \theta_1 \beta_1)}{\sqrt{1 + (A_2 \sigma_{\epsilon_1})^2} \sqrt{1 + A_4^2 \sigma_U^2}}$, $A_6 = \frac{c(\theta_1 \beta_2)}{\sqrt{1 + (A_2 \sigma_{\epsilon_1})^2} \sqrt{1 + A_4^2 \sigma_U^2}}$ and $A_7 = A_5^2/(A_6^2 + 1)$. Similarly $E[RUx] = E_x[xE[RU|x]]$ and

$$
\begin{aligned}
E_x[xE[RU|x]] &= \int x \frac{\sigma_U^2 A_4}{\sqrt{1 + A_4^2 \sigma_U^2}} \phi[A_5 + A_6 x] \phi(x; 0, \sigma_x^2) \mathrm{d}x \\
&= \frac{\sigma_U^2 A_4}{\sqrt{1 + A_4^2 \sigma_U^2}} \int x \phi[A_5 + A_6 x] \phi(x; 0, \sigma_x^2) \mathrm{d}x \\
E[RUx] &= \frac{e^{-\frac{1}{2} A_7}}{\sqrt{2\pi} \sqrt{A_6^2 + 1}} A_8
\end{aligned}
$$

where $A_8 = -A_6 A_5/(A_6^2 + 1)$.

Hence we get $E[RY_2] = \beta_3 E[R] + \beta_4 E[Rx] + E[RU]$ and $E[RY_2 x] = \beta_3 E[Rx] + \beta_4 E[Rx^2] + E[RUx]$, and we have all terms needed for the least false calculation.

5.5 Performance under MNAR Dropout

We found $E[R], E[Rx], E[Rx^2], E[R\epsilon_2]$ and $E[R\epsilon_2 x]$ in Chapter 3 in equations (3.47), (3.48), (3.49), (3.50) and (3.51) respectively. Thus we still need to find $E[RU]$ and $E[RUx]$.

We start with $E[RU] = E_x[E[RU|x]]$ as previously. This time

$$
\begin{aligned}
E[RU|x] &= \int \int U \operatorname{expit} \{\theta_0 + \theta_2 Y_2\} \phi(U; 0, \sigma_U^2) \phi(\epsilon_2; 0, \sigma_{\epsilon_2}^2) \mathrm{d}U \, \mathrm{d}\epsilon_2 \\
&= \frac{\sigma_U^2 B_4}{\sqrt{1 + B_4^2 \sigma_U^2}} \phi\left(\frac{B_3}{\sqrt{1 + B_4^2 \sigma_U^2}}\right)
\end{aligned}
$$

where $B_1 = c(\theta_0 + \theta_2 \beta_3 + \beta_4 x) + c\theta_2 U$, $B_2 = c\theta_2$, $B_3 = \frac{c(\theta_0 + \theta_2(\beta_3 + \beta_4 x))}{\sqrt{1 + (B_2 \sigma_{\epsilon_2})^2}}$ and $B_4 = \frac{c\theta_2}{\sqrt{1 + (B_2 \sigma_{\epsilon_2})^2}}$.

Then

$$
\begin{aligned}
E[RU] &= E_x[E[RU|x]] \\
&= \frac{\sigma_U^2 B_4}{\sqrt{1 + B_4^2 \sigma_U^2}} \frac{e^{-\frac{1}{2}B_7}}{\sqrt{2\pi}\sqrt{B_6^2 + 1}}
\end{aligned}
$$

where $B_7 = B_5^2/(B_6^2 + 1)$.
Similarly, $E[RUx] = E_x[xE[RU|x]]$ and

$$
\begin{aligned}
E_x[xE[RU|x]] &= \int x \frac{\sigma_U^2 B_4}{\sqrt{1 + B_4^2 \sigma_U^2}} \phi[B_5 + B_6 x]\phi(x; 0, \sigma_x^2)\mathrm{d}x \\
&= \frac{\sigma_U^2 B_4}{\sqrt{1 + B_4^2 \sigma_U^2}} \int x\phi[B_5 + B_6 x]\phi(x; 0, \sigma_x^2)\mathrm{d}x \\
E[RUx] &= \frac{e^{-\frac{1}{2}B_7}}{\sqrt{2\pi}\sqrt{B_6^2 + 1}} B_8
\end{aligned}
$$

where $B_8 = -B_6 B_5/(B_6^2 + 1)$.

Hence we get $E[RY_2] = \beta_3 E[R] + \beta_4 E[Rx] + E[RU] + E[R\epsilon_2]$ and $E[RY_2 x] = \beta_3 E[Rx] + \beta_4 E[Rx^2] + E[RUx] + E[Rx\epsilon_2]$. We will now investigate these results numerically.

5.6   Simulations and Numerical Investigation

We will do this numerical investigation in two parts as in the previous chapters. In the first part we want to check that the theory and our least false values are correct by comparing the expected values with simulated ones. In the second part we explore in detail how the misspecification affects the β estimates, using the theoretical results. Now that we have working formulas there is no need for simulations here.

As before we generated a scalar $N(0,1)$ variable x, then we generated the longitudinal means $\mu_1 = \beta_1 + \beta_2 x$, $\mu_2 = \beta_3 + \beta_4 x$. This was followed by (Y_1, Y_2) from a bivariate normal distribution with mean (μ_1, μ_2) using a shared effect U to generate the correlation. In all of the following simulations, unless it is stated otherwise, we take the same parameters used in previous chapter. Our choice of parameters will keep the dropout always about 50%.

5.6.1   The limiting values compared to the simulated values for different sample size

Here we attempt to show what happens as sample size increases, i.e. at what n can we be confident that the theory is reasonable. One can note from Table 5.1 as we expected as n increases the simulated and theoretical values become close to each other. Note that $\hat{\beta}_3$ and $\hat{\beta}_4$ are the estimated values produced by the simulations when we fit MAR to data that are MNAR or SP.

Results at n=100000 show that the simulated values all match the theoretical approximations very well, which support the use of the theory. Comparing the results here with the corresponding results in Chapter 3, we note that here we need a larger sample size to achieve n that gives us an acceptable result. This may be related to the extra variability of CC observed in Chapter 2, perhaps because we use fewer observations in estimation.

Based on the $n = 100000$ result, Table 5.2 gives detail on the various components of β_3^* and β_4^*. We see that all dropout methods lead to bias, of generally similar magnitude, but the causes (the components) can differ considerably.

Table 5.1: The limiting values under different sample sizes

n		SP	MAR	MNAR
1000	$\hat{\beta}_3$	-0.82	-0.89	-0.78
	β_3^*	-0.84	-0.90	-0.77
	$\hat{\beta}_4$	-1.11	-0.95	-0.95
	β_4^*	-1.00	-0.97	-0.95
10000	$\hat{\beta}_3$	-0.82	-0.88	-0.76
	β_3^*	-0.84	-0.90	-0.77
	$\hat{\beta}_4$	-1.00	-0.95	-0.94
	β_4^*	-1.00	-0.97	-0.95
50000	$\hat{\beta}_3$	-0.84	-0.88	-0.76
	β_3^*	-0.84	-0.90	-0.77
	$\hat{\beta}_4$	-0.99	-0.95	-0.96
	β_4^*	-1.00	-0.97	-0.95
100000	$\hat{\beta}_3$	-0.84	-0.89	-0.77
	β_3^*	-0.84	-0.90	-0.77
	$\hat{\beta}_4$	-1.00	-0.95	-0.95
	β_4^*	-1.00	-0.97	-0.95

Table 5.2: Comparison of simulated and least false values under $n = 100000$

	SP	MAR	MNAR
$\bar{R}$	0.50	0.50	0.50
$E[R]$	0.50	0.50	0.50
$\bar{R}x$	-0.00	-0.20	-0.11
$E[Rx]$	0.00	-0.20	-0.11
$\bar{R}x^2$	0.50	0.50	0.50
$E[Rx^2]$	0.50	0.50	0.50
$\bar{R}Y_2$	-0.42	-0.25	-0.28
$E[RY_2]$	-0.42	-0.25	-0.28
$\bar{R}Y_2x$	-0.50	-0.30	-0.39
$E[RY_2x]$	-0.50	-0.30	-0.39
$\hat{\beta}_3$	-0.83	-0.88	-0.76
β_3^*	-0.84	-0.90	-0.77
$\hat{\beta}_4$	-1.00	-0.95	-0.95
β_4^*	-1.00	-0.97	-0.95

5.6.2 The 95% nominal CI for the limiting values

Here we will show that for each limiting value calculated by the expectation, the simulated values (SV) are within noise of the theoretical values (TV) for large sample sizes (n=1000). We estimate the noise from the simulations; that is we get a confidence interval from the simulations and reassurance that the population values are within these. We simulate 100 samples of size 1000, then find the $\hat{\beta}$ for each sample then calculate the reference interval instead of standard confidence interval, so we will say: CI=$\bar{\hat{\beta}} \pm 2\,SD(\beta)$. We did this for the SP, MAR and MNAR dropout models, each with just one set of dropout parameters and all with the missingness percentage at about 50% dropout. In Tables 5.3 and 5.4, the theoretical values are all being within the corresponding reference intervals.

Table 5.3: The 95% reference interval for β_3^* using 100 samples of size 1000.

	TV	SV	Lower Bound	Upper Bound
MAR	-0.8842	-0.8852	-0.9188	-0.8516
MNAR	-0.7727	-0.7626	-0.7900	-0.7351
SP	-0.8408	-0.8321	-0.8590	-0.8053

5.6.3 The effect of dropout on the limiting values β_3^* and β_4^*

We now explore the effect of different dropout parameters under fixed β and $\sigma_x^2 = 1$, $\sigma_U = \sigma_{\epsilon_1} = \sigma_{\epsilon_2} = \sqrt{0.5}$. We choose values that keep the dropout rate close to 50%, and use θ^{sp}, θ^M and θ^{MN} to denote values under SP, MAR and MNAR respectively. Recall

Table 5.4: The 95% reference interval for β_4^* using 100 samples of size 1000.

	TV	SV	Lower Bound	Upper Bound
MAR	-0.9546	-0.9536	-0.9877	-0.9194
MNAR	-0.9508	-0.9475	-0.9782	-0.9167
SP	-1.0000	-0.9996	-1.0249	-0.9743

that the true values are $\beta_3 = \beta_4 = -1$.

In Table 5.5, we show β_3^* and β_4^* by solving the equations given in Section 5.3 for the given values of θ_0^{sp} and θ_1^{sp} under SP, with the same parameters used to generate the logistic model as mentioned before. These values of θ all give a dropout rate of around 50%.

Under SP, the least false values of β_4^* are very close to the true quantity. Hence we only investigate the effect of θ's on β_3. The bias in β_3 is larger as the value of θ_1^{sp} get bigger. For negative θ_1^{sp}, we note that β_3 tends to be smaller than its true value. So the chance of dropout goes up with Y_2, so we lose the large Y_2 values and the mean of the observed data at time 2 is too low. So the intercept term (β_3) is too low. The opposite happens for positive θ_1^{sp}.

In Table 5.6, we show β_3^* and β_4^* by solving the equations given in Section 5.4 for the given values of θ_0^M and θ_1^M under MAR, with the same parameters used to generate the logistic model as mentioned before. These values of θ all give a dropout rate of around 50%. Similar conclusions to Table 5.5 are also drawn here for β_3^*. For β_4^*, there is now bias, which decreases as the negative θ_1^M increases, while for positive θ_1^M the bias increases as the positive θ_1^M increases.

Turning to the next table, under MNAR, Table 5.7, we show β_3^* and β_4^* by solving the equations given in Section 5.5 for the given values of θ_0^{MN} and θ_2^{MN} under MNAR, with the same parameters used to generate the logistic model as mentioned before. These values of θ all give a dropout rate of around 50%. Similar conclusions to Table 5.6 are also drawn here for both β_3^* and β_4^*. In Table 5.7 β_4 is biased towards zero as previously discussed (Appendix A.6). But β_3 is biased either down or up depending upon whether dropout is associated with large Y_2 (negative θ_2^{MN}), or small Y_2 (positive θ_2^{MN}). If dropout is associated with small Y_2, then the mean in the observed data will be too high. The opposite is true if dropout is associated with large Y_2.

Table 5.5: Effect on dropout parameters of fitting the CC method under SP. True $\beta_3 = -1$ and β_4=-1.

θ_0^{sp}	θ_1^{sp}	β_3^*	β_4^*
0	$-\sqrt{2}$	-1.2860	-1.0000
0	$-\sqrt{0.5}$	-1.1592	-1.0000
0	$\sqrt{0.5}$	-0.8408	-1.0000
0	$\sqrt{2}$	-0.7140	-1.0000

Table 5.6: Effect on dropout parameters of fitting the CC method under MAR. True $\beta_3 = -1$ and β_4=-1.

θ_0^M	θ_1^M	β_3^*	β_4^*
-2	-1	-1.2097	-0.8809
-1	-0.5	-1.1158	-0.9546
1	0.5	-0.8842	-0.9546
2	1	-0.7903	-0.8809

Table 5.7: Effect on dropout parameters of fitting the CC method under MNAR. True $\beta_3 = -1$ and β_4=-1.

θ_0^{MN}	θ_2^{MN}	β_3^*	β_4^*
-1	-1	-1.4147	-0.8504
-0.5	-0.5	-1.2273	-0.9508
0.5	0.5	-0.7727	-0.9508
1	1	-0.5853	-0.8504

5.6.4 The effect of the variance of the random effect on the limiting values β_3^* and β_4^*

In the following tables we will study the effect of changing the variance of the random effect and measurement error on the limiting values at different dropout models. We take $\sigma_U^2 = \sigma_{\epsilon_1}^2 = \sigma_{\epsilon_2}^2$. Each time, we keep the dropout percentage at about 50%.

Table 5.8: The effect of the variance σ_U^2 on the limiting values β_3^* and β_4^*. True $\beta_3 = -1$ and β_4=-1.

		SP	MAR	MNAR
$\sigma_U = \sqrt{0.5}$	β_3^*	-0.8408	-0.8965	-0.7727
	β_4^*	-1.0000	-0.9665	-0.9508
$\sigma_U = 1$	β_3^*	-0.6936	-0.7763	-0.5628
	β_4^*	-1.0000	-0.9142	-0.9086
$\sigma_U = \sqrt{3}$	β_3^*	-0.1923	-0.4117	0.1505
	β_4^*	-1.0000	-0.7964	-0.7870
$\sigma_U = \sqrt{6}$	β_3^*	0.3946	0.0076	0.9832
	β_4^*	-1.0000	-0.6938	-0.6807
$\sigma_U = \sqrt{12}$	β_3^*	1.2705	0.6240	2.2229
	β_4^*	-1.0000	-0.5841	-0.5748

Changing the variance parameters and keeping the dropout percentage at about 50% has a remarkable effect on the bias. From Table 5.8, it is clear that as σ_U (and consequentially σ_{ϵ_1} and σ_{ϵ_2}) increases, the limiting values β_3^* goes further from the true value ($\beta_3^G = -1, \beta_4^G = -1$) and hence the absolute bias increases, which means that the large error variances imply poor results. In short we can conclude that the more variability, the more bias.

5.6.5 Contour plots

Here we want to show how the limiting values β^* change when we change the dropout parameter θ. For shared parameter, we produced contour plots of β^* for a grid with $-1.5 < \theta_0^{sp} < 1.5$ and $-1.5 < \theta_1^{sp} < 1.5$ for each dropout model. Figures 5.1 - 5.5 show the results. We do not present a plot for β_4^* under SP, because the CC method gives close estimates for β_4 and no real bias.

In Figure 5.1, β_3^* under SP can take a value from about -1.4 to -0.6 for the θ in the aforementioned grid. The white area in the plot corresponds to the correct value which is $\beta_3^G = -1$. We can conclude that from the contour plot of β_3^* under SP, that in order to minimise the bias in β_3^*, we should choose θ_1 to be around zero, which is equivalent to MCAR. The biggest change is in the vertical direction, for example at $\theta_1 = 0$, then $\beta_3^* = -1$ as expected. For negative θ_1, the dropout is associated with large Y_2, so Y_2 tends to be low. Hence β_3^* is lower than what it should be. The opposite happens for a positive θ_1.

In Figure 5.2, β_3^* under MAR can take a value from about -1.8 to -0.8 for the chosen range of θ. For negative θ_1, β_3^* get closer to the true value (-1) as θ_0 increases. However, the bias for β_3^* is greater at negative θ_1 than at positive θ_1. At low θ_1 then β_3^* is very low for low θ_0 but high for high θ_0. In contrast at high θ_1, β_3^* is high for all θ_0.

In Figure 5.3, β_3^* under MNAR can take a value from about -2.0 to -0.5 for the chosen range of θ. Again the bias for β_3^* is greater at negative θ_2 than at positive θ_2 as happens in Figure 5.2.

In Figure 5.4, β_4^* under MAR can take a value from about -1.4 to -0.2 for the θs changes in aforementioned grid. For positive θ_1, β_4^* get closer to the true value (-1) as θ_0 increases. However, the bias for β_4^* is greater at positive θ_1 than at negative θ_1. At low θ_1 then β_4^* is very low for high θ_0 but high for low θ_0. In contrast at high θ_1, β_4^* is low for all θ_0.

In Figure 5.5, β_4^* under MNAR can take a value from about -1.2 to -0.4 for the θs changes in aforementioned grid. The interpretation is as for the previous Figure.

In conclusion, the dropout model MAR for β_3^* has less bias than the others, while the dropout model MNAR for β_4^* has more bias than MAR. This is because when we compare Figure 5.2 with Figures 5.1 and 5.3, we find that the bias area in MAR is smaller than the bias area in SP and MNAR. Similarly, when comparing Figure 5.4 with Figure 5.5 the bias area in MAR is larger than the bias area in MNAR. Results from the contour plots support our conclusions from Table 5.5, Table 5.6 and Table 5.7.

Figure 5.1: Contour plot of β_3^* under SP.
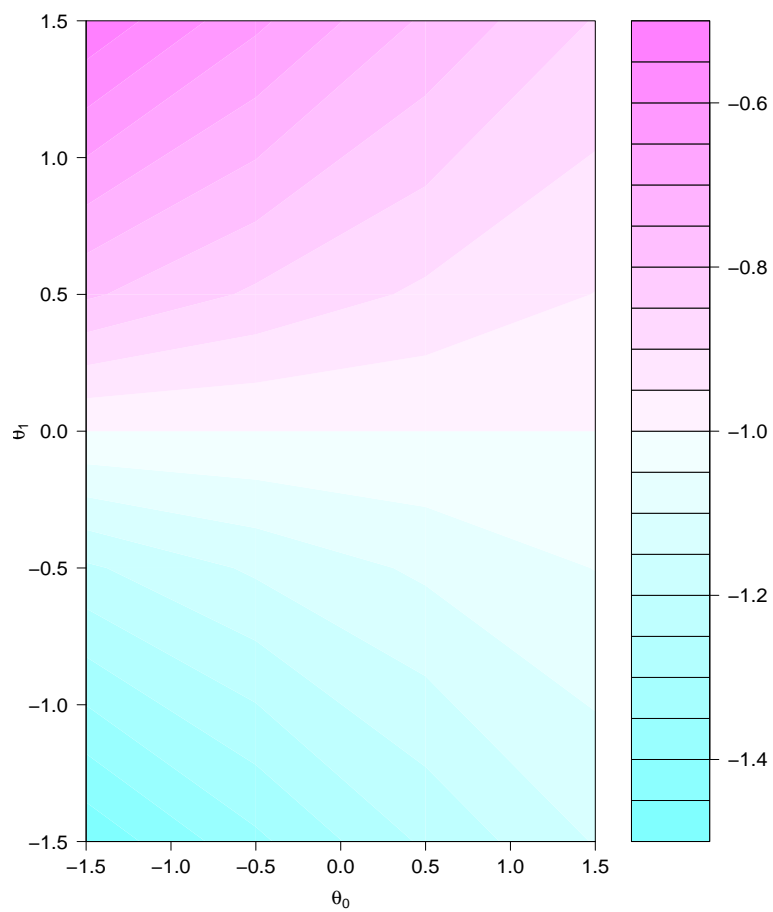
Figure 5.2: Contour plot of β_3^* under MAR.
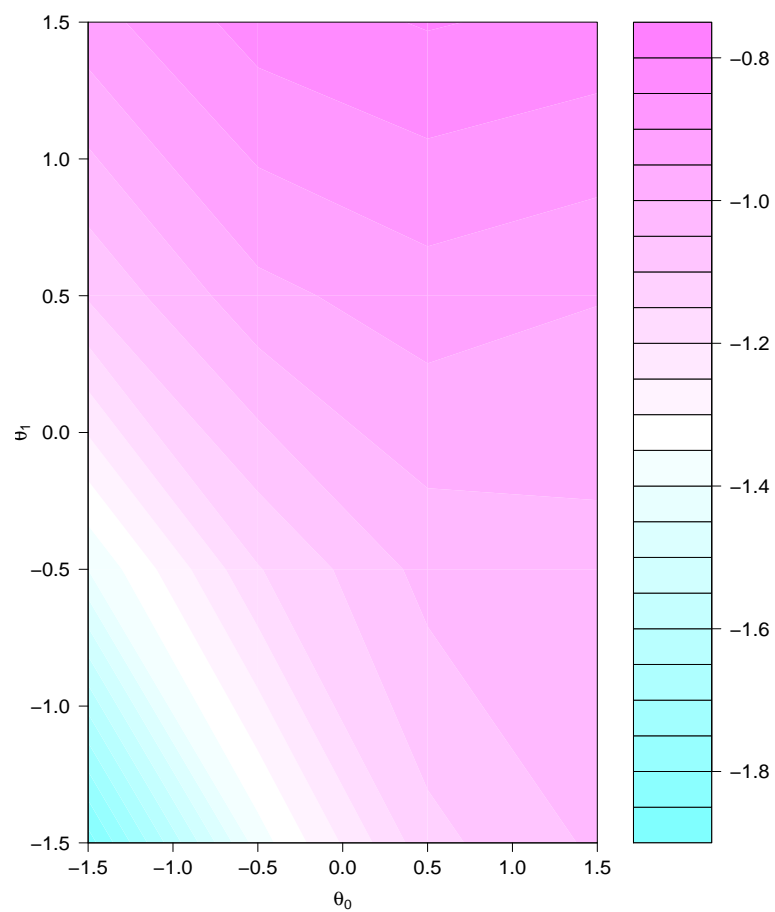
Figure 5.3: Contour plot of β_3^* under MNAR.
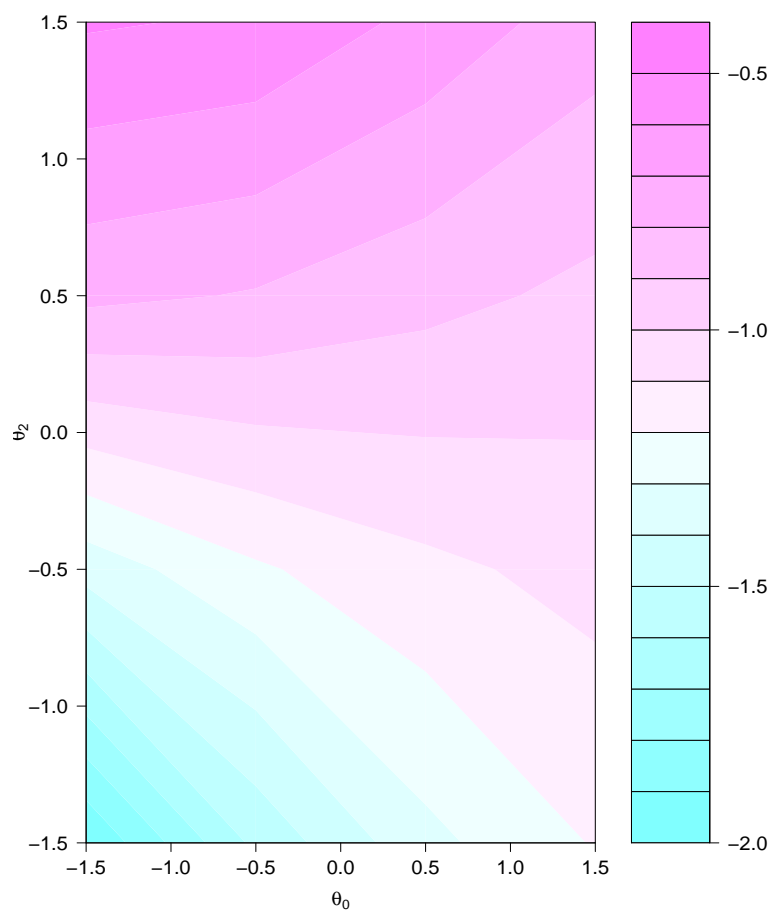
Figure 5.4: Contour plot of β_4^* under MAR.
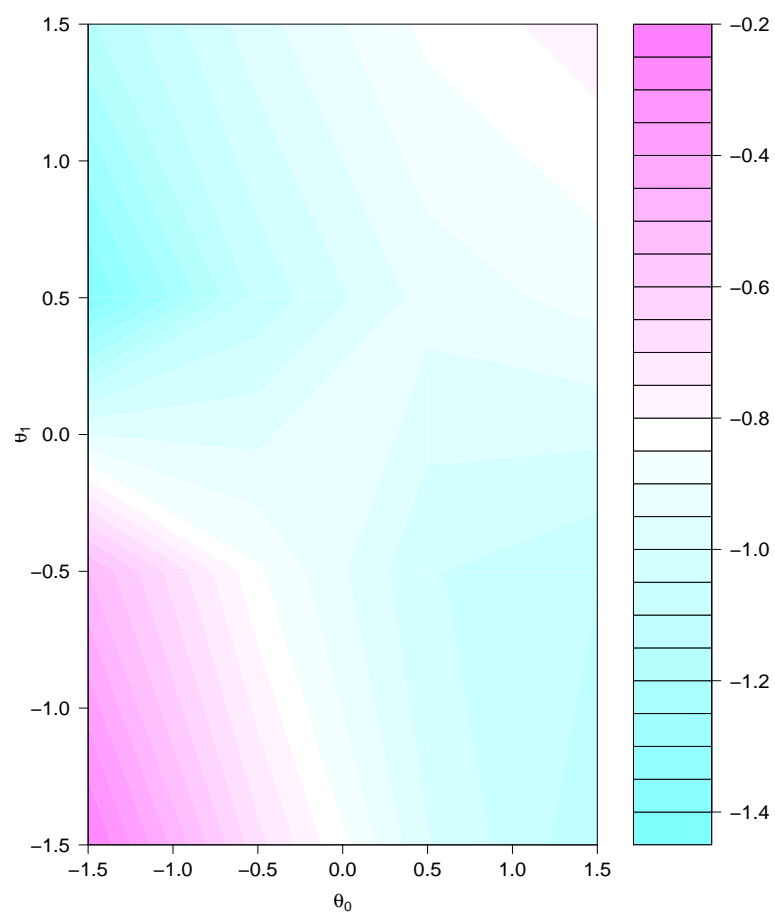
Figure 5.5: Contour plot of β_4^* under MNAR.
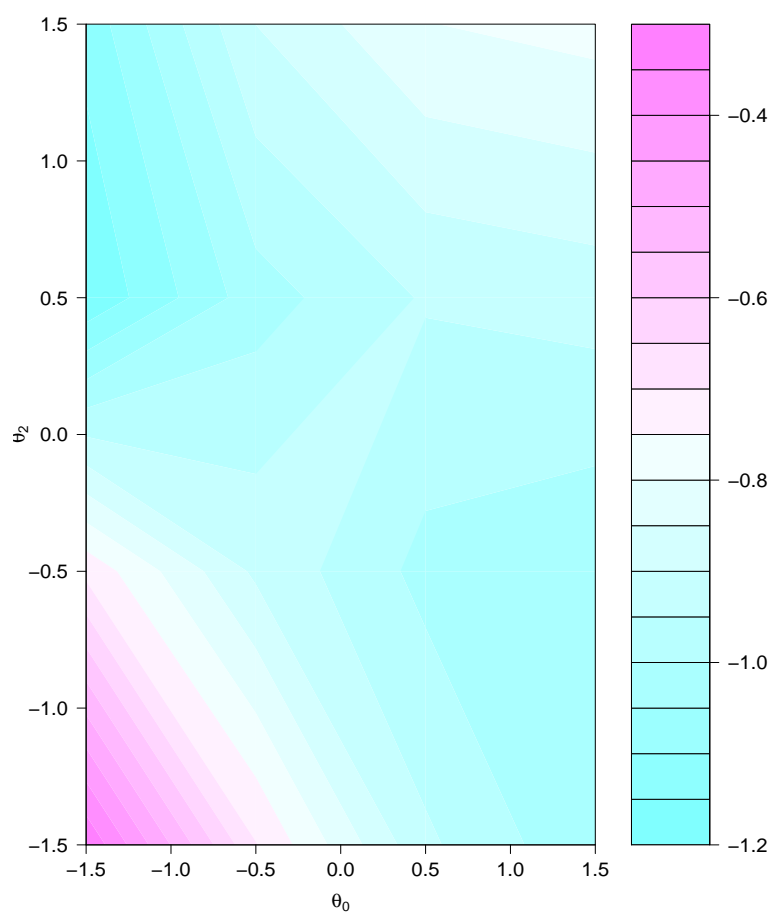
5.7   Discussion and Conclusion

In this chapter we calculated the least false regression values under a complete case analysis. We compared these expected values under different dropout models to the generating values β_3^G and β_4^G. We calculated the bias in the expected parameters in each dropout model. It is clear that the CC does not seem to work under any dropout model as shown in Tables 5.5 - 5.8 and Figures 5.1 - 5.5; since $\beta_3^* \neq \beta_3^G$ and $\beta_4^* \neq \beta_4^G$, or, in other words, there is bias.

It is interesting to note one exception that the CC method gives consistent estimates for β_4 under SP dropout model. This is straightforward. We will turn in the next chapter to a further analysis of least false values.

Chapter 6

Performance of The Linear Mixed Effect Method Under Shared Parameter, MAR and MNAR Dropout

6.1 Introduction

In this chapter we investigate the linear mixed effect method (LME) which is based on a maximum likelihood estimating approach. Following the methods of the previous chapters, we take the two timepoints situation as a special case and investigate how the LME method performs under shared parameter, MAR and MNAR dropout. Derivation and illustration of theoretical least false values are made under SP, MAR and MNAR dropout. To check, we compare the theoretical least false values with simulation results for selected parameter combinations.

6.2 Linear Mixed Effect Method with Two Timepoints

Maximum likelihood estimation gives consistent estimators under MAR as we found in Chapter 2. Assuming a Gaussian random intercept model, the score equation of current interest is:

$$\sum_{i=1}^{n} \left[R_i \left\{ X_i^T V^{-1}(Y_i - X_i\hat{\beta}) \right\} + \frac{(1 - R_i)}{\sigma_1^2} \left\{ x_{i1}(Y_{i1} - x_{i1}^T\hat{\beta}) \right\} \right] = 0 \qquad (6.1)$$

where $Y_i = (Y_{i1}, Y_{i2})$, X_i is a 2×4 design matrix associated with subject i which is $X_i = \begin{pmatrix} 1 & x_i & 0 & 0 \\ 0 & 0 & 1 & x_i \end{pmatrix}$ and we have used x_{i1}^T as notation for the first row of X_i, thus

$x_{i1}^T = (1, x_i, 0, 0)$, $\hat{\beta}^T = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4)$, and $V = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, where σ_1, σ_2 and ρ are as defined in the Introduction. We can re arrange the terms in (6.1) to be:

$$\sum_{i=1}^n \left[R_i \left\{ X_i^T V^{-1} X_i \right\} + \frac{(1 - R_i)}{\sigma_1^2} \left\{ x_{i1} x_{i1}^T \right\} \right] \hat{\beta} = \sum_{i=1}^n \left[R_i \left\{ X_i^T V^{-1} Y_i \right\} + \frac{(1 - R_i)}{\sigma_1^2} \left\{ x_{i1} Y_{i1} \right\} \right]. \tag{6.2}$$

These components are in detail:

$$V^{-1} = K \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} \tag{6.3}$$

where $K = 1/\sigma_1^2 \sigma_2^2 (1 - \rho^2)$, and

$$X_i^T V^{-1} X_i = K \begin{pmatrix} \sigma_2^2 & \sigma_2^2 x_i & -\rho\sigma_1\sigma_2 & -\rho\sigma_1\sigma_2 x_i \\ \sigma_2^2 x_i & \sigma_2^2 x_i^2 & -\rho\sigma_1\sigma_2 x_i & -\rho\sigma_1\sigma_2 x_i^2 \\ -\rho\sigma_1\sigma_2 & -\rho\sigma_1\sigma_2 x_i & \sigma_1^2 & \sigma_1^2 x_i \\ -\rho\sigma_1\sigma_2 x_i & -\rho\sigma_1\sigma_2 x_i^2 & \sigma_1^2 x_i & \sigma_1^2 x_i^2 \end{pmatrix}. \tag{6.4}$$

Also

$$x_{i1} x_{i1}^T = \begin{pmatrix} 1 & x_i & 0 & 0 \\ x_i & x_i^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{6.5}$$

Similarly for the right hand side of equation (6.2)

$$X_i^T V^{-1} Y_i = K \begin{pmatrix} \sigma_2^2 Y_{i1} - \rho\sigma_1\sigma_2 Y_{i2} \\ \sigma_2^2 Y_{i1} x_i - \rho\sigma_1\sigma_2 Y_{i2} x_i \\ \sigma_1^2 Y_{i2} - \rho\sigma_1\sigma_2 Y_{i1} \\ \sigma_1^2 Y_{i2} x_i - \rho\sigma_1\sigma_2 Y_{i1} x_i \end{pmatrix}. \tag{6.6}$$

Finally

$$x_{i1} Y_{i1} = \begin{pmatrix} Y_{i1} \\ x_i Y_{i1} \\ 0 \\ 0 \end{pmatrix}. \tag{6.7}$$

We can collect all of these, divide by n, and use the law of large numbers to replace the summation by the expectations as the following:

$$E\left[R\left\{ X^T V^{-1} X \right\} + \frac{(1 - R)}{\sigma_1^2} \left\{ x_1^T x_1 \right\} \right] \beta^* = E\left[R\left\{ X^T V^{-1} Y \right\} + \frac{(1 - R)}{\sigma_1^2} \left\{ x_1 Y_1 \right\} \right] \tag{6.8}$$

In the left hand side of equation (6.8) we will have two parts. First

$$
E\left[R\left\{X^T V^{-1} X\right\}\right] = K \begin{pmatrix} \sigma_2^2 E[R] & \sigma_2^2 E[Rx] & -\rho\sigma_1\sigma_2 E[R] & -\rho\sigma_1\sigma_2 E[Rx] \\ \sigma_2^2 E[Rx] & \sigma_2^2 E[Rx^2] & -\rho\sigma_1\sigma_2 E[Rx] & -\rho\sigma_1\sigma_2 E[Rx^2] \\ -\rho\sigma_1\sigma_2 E[R] & -\rho\sigma_1\sigma_2 E[Rx] & \sigma_1^2 E[R] & \sigma_1^2 E[Rx] \\ -\rho\sigma_1\sigma_2 E[Rx] & -\rho\sigma_1\sigma_2 E[Rx^2] & \sigma_1^2 E[Rx] & \sigma_1^2 E[Rx^2] \end{pmatrix},
$$

$$(6.9)$$

and second:

$$
\frac{(1-R)}{\sigma_1^2} x_1 x_1^T = \frac{1}{\sigma_1^2} \begin{pmatrix} 1 - E[R] & E[x] - E[Rx] & 0 & 0 \\ E[x] - E[Rx] & E[x^2] - E[Rx^2] & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{6.10}
$$

Similarly, the right hand side is

$$
K \begin{pmatrix} \sigma_2^2 E[RY_1] - \rho\sigma_1\sigma_2 E[RY_2] \\ \sigma_2^2 E[RY_1 x] - \rho\sigma_1\sigma_2 E[RY_2 x] \\ \sigma_1^2 E[RY_2] - \rho\sigma_1\sigma_2 E[RY_1] \\ \sigma_1^2 E[RY_2 x] - \rho\sigma_1\sigma_2 E[RY_1 x] \end{pmatrix} + \frac{1}{\sigma_1^2} \begin{pmatrix} E[Y_1] - E[RY_1] \\ E[Y_1 x] - E[RY_1 x] \\ 0 \\ 0 \end{pmatrix}. \tag{6.11}
$$

Finally to find the least false value β^* we take the inverse of the matrix in the left hand side of equation (6.8) and multiply this inverse by the matrix in the right hand side, this will yield the array of the least false values $\beta^{*T} = (\beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*)$. Expressions of $E[R]$, $E[Rx]$, $E[Rx^2]$, $E[RY_1]$, and $E[RY_2]$ have been obtained in the previous chapters under our different dropout models. In the following section we will show in simulations how the LME method performs under MAR, SP and MNAR dropout model using a similar technique.

6.3 Simulations and Numerical Investigation

We will do this numerical investigation in two parts as in the previous chapters. In the first part we want to check that the theory and our least false values are correct by comparing the expected values with simulated ones. In the second part we explore in detail how the misspecification affects the β estimates, using the theoretical results. Now that we have working formulas there is no need for simulations here. In all of the following simulations, unless it is stated otherwise, we take the same parameters used in previous chapter. Our choice of parameters will keep the dropout always about 50%. So we have $\beta = (-2, -2, -1, -1)$, $\sigma_1^2 = 0.5$, $\sigma_2^2 = 0.5$, $\sigma_x^2 = 1$, $\rho = 0.5$.

6.3.1 The limiting values compared to the simulated values for different sample size

In Table 6.1 we attempt to show what happens as sample size increases, i.e. at what n can we be confident that the theory is reasonable. One can note as expected that as n increases the simulated and theoretical values become close to each other.

Results at n=100000 show that the simulated values all match the theoretical approximations very well, which support the use of the theory. Comparing the result here with the corresponding result in Chapter 3, we note that here we need a larger sample size to achieve n that gives us an acceptable result.

Table 6.1: The limiting values β_3^* and β_4^* under different sample sizes

n		MAR	SP	MNAR
1000	$\hat{\beta}_3$	-0.95	-0.93	-0.86
	$\hat{\beta}_4$	-0.99	-1.00	-0.94
	β_3^*	-1.00	-0.92	-0.83
	β_4^*	-1.00	-1.00	-0.96
10000	$\hat{\beta}_3$	-1.00	-0.91	-0.81
	$\hat{\beta}_4$	-1.02	-1.00	-0.96
	β_3^*	-1.00	-0.92	-0.83
	β_4^*	-1.00	-1.00	-0.96
50000	$\hat{\beta}_3$	-0.99	-0.91	-0.82
	$\hat{\beta}_4$	-1.00	-1.00	-0.95
	β_3^*	-1.00	-0.92	-0.83
	β_4^*	-1.00	-1.00	-0.96
100000	$\hat{\beta}_3$	-1.00	-0.92	-0.82
	$\hat{\beta}_4$	-1.00	-1.00	-0.96
	β_3^*	-1.00	-0.92	-0.83
	β_4^*	-1.00	-1.00	-0.96

6.3.2 The 95% nominal CI for the limiting values

Here we will show that for each limiting value calculated by the expectation, the simulated values (SV) are within noise of the theoretical values (TV) for large sample sizes using the methods of previous chapters. Tables 6.2 and 6.3 show that the theoretical least false values are all within the reference ranges for each dropout model. The MAR model is included for completeness.

Table 6.2: The 95% reference interval for β_3^* using 100 samples of size 1000.

	TV	SV	Lower	Upper
MAR	-1.0000	-0.9995	-1.0353	-0.9637
MNAR	-0.8295	-0.8227	-0.8331	-0.8124
SP	-0.9204	-0.9209	-0.9430	-0.8987

Table 6.3: The 95% reference interval for β_4^* using 100 samples of size 1000.

	TV	SV	Lower	Upper
MAR	-1.0000	-1.0004	-1.0659	-0.9350
MNAR	-0.9631	-0.9581	-1.0164	-0.8997
SP	-1.0000	-0.9995	-1.0585	-0.9405

6.3.3 The effect of dropout probabilities on the limiting values β_3^* and β_4^*

We explore the effect of different dropout parameters under fixed $\beta = (-2, -2, -1, -1)$, $\sigma_x^2 = 1$, $\rho = 0.5$ and $\sigma_1 = \sigma_2 = \sqrt{0.5}$. We choose values that keep the dropout rate close to 50%, and use θ^{sp}, θ^M and θ^{MN} to denote values under SP, MAR and MNAR respectively.

As the LME works under MAR, we do not show the table for how the least false values β_3^* and β_4^* are affected by changing the dropout probabilities under MAR, but we show the tables under SP and MNAR.

In Tables 6.4 and 6.5, we show β_3^* and β_4^* by solving the equations given in Sections 6.2 and 6.3 for the given values of θ_0^{sp} and θ_1^{sp} under SP and θ_0^{MN} and θ_2^{MN} under MNAR, with the same parameters used to generate the logistic model as mentioned before. These values of θ all give a dropout rate of around 50%. Recall that the true values are $\beta_3^* = \beta_4^* = -1$. We see that LME seems to work for β_4^* under the SP model because there is no bias for β_4^* in Table 6.4. Recall that β_4^* was also almost unbiased under IPW (Chapter 4, Table 4.5) and CC (Chapter 5, Table 5.5).

Table 6.4: Effect of the dropout parameters $(\theta_0^{sp}, \theta_1^{sp})$ on fitting the LME method under SP. True $\beta_3 = -1$ and β_4=-1.

θ_0^{sp}	θ_1^{sp}	β_3^*	β_4^*
0	$\sqrt{0.5}$	-0.9204	-1.0000
0	$-\sqrt{0.5}$	-1.0796	-1.0000
0	$\sqrt{2}$	-0.8570	-1.0000
0	$-\sqrt{2}$	-1.1430	-1.0000

Table 6.5: Effect of the dropout parameters $(\theta_0^{MN}, \theta_2^{MN})$ on fitting the LME method under MNAR. True $\beta_3 = -1$ and β_4=-1.

θ_0^{MN}	θ_2^{MN}	β_3^*	β_4^*
0.5	0.5	-0.8295	-0.9631
-0.5	-0.5	-1.1705	-0.9631
1	1	-0.6889	-0.8878
-1	-1	-1.3111	-0.8878

6.3.4 Contour plots

Here we want to show in plots how the limiting values β_3^* and β_4^* change when we change the dropout probabilities. We take the dropout probabilities in the range from -1.5 to 1.5. We do not attach the contour plot for β_4^* under SP as there is no bias, which gives the value of β_4^* always equals -1 at any θ_0 and θ_1 in the range (-1.5,1.5). Also, we do not show the plots under MAR since the LME works under MAR.

We can conclude that from the contour plot of β_3^* under SP (Figure 6.1), that in order to minimise the bias in β_3^*, we should choose θ_1 to be around zero, which is equivalent to MCAR. For negative θ_1, the dropout is associated with large U, so Y_1 and Y_2 both tend to be low. Hence β_3^* is lower than what it should be. The opposite happens for a positive θ_1. The same comments apply to Figure 6.2 which show the contour plot of β_3^* under MNAR.

Figure 6.3 show the contour plot of β_4^* under MNAR. Here we get similar result to Figure 3.2 in Chapter 3. Again we get attenuation as θ_2 moves away from zero in either direction.
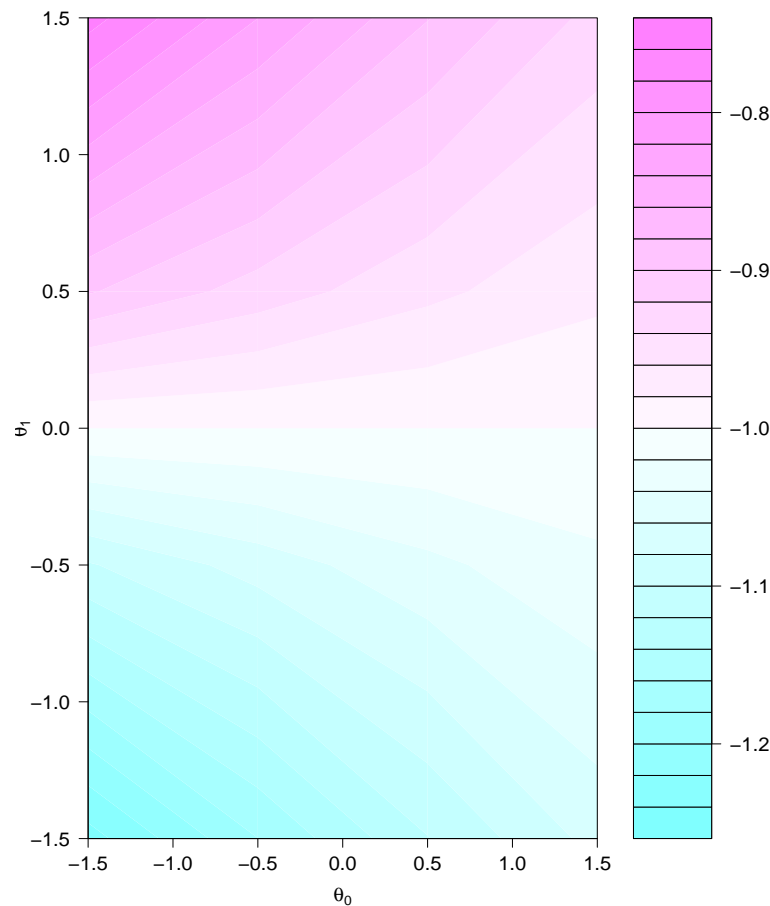
Figure 6.1: Contour plot of β_3^* under SP.

Figure 6.2: Contour plot of β_3^* under MNAR.
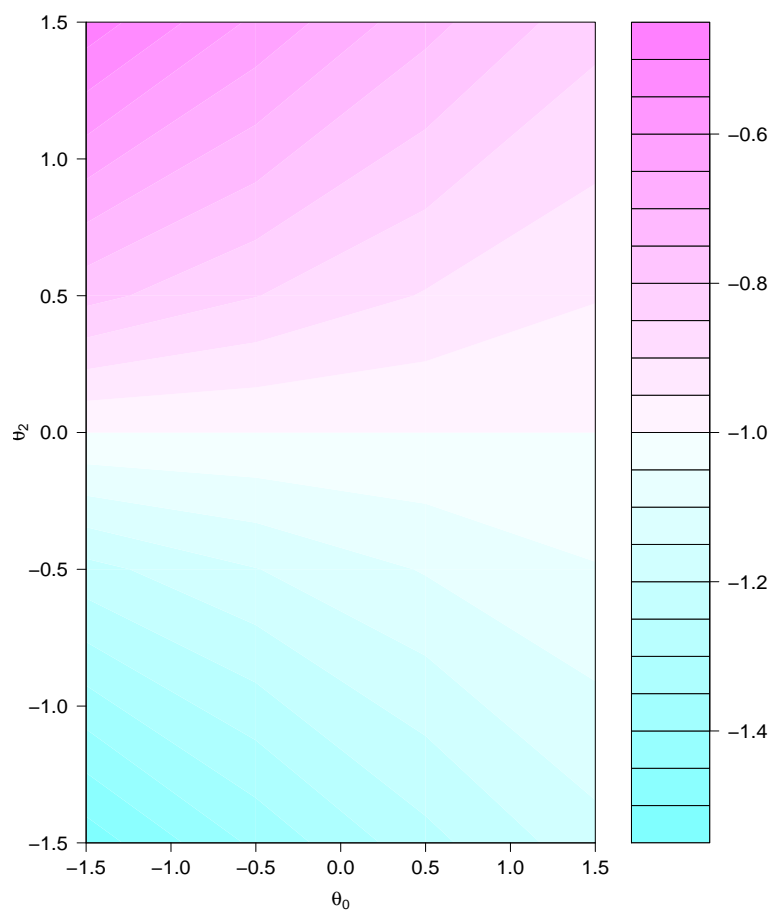
Figure 6.3: Contour plot of β_4^* under MNAR.
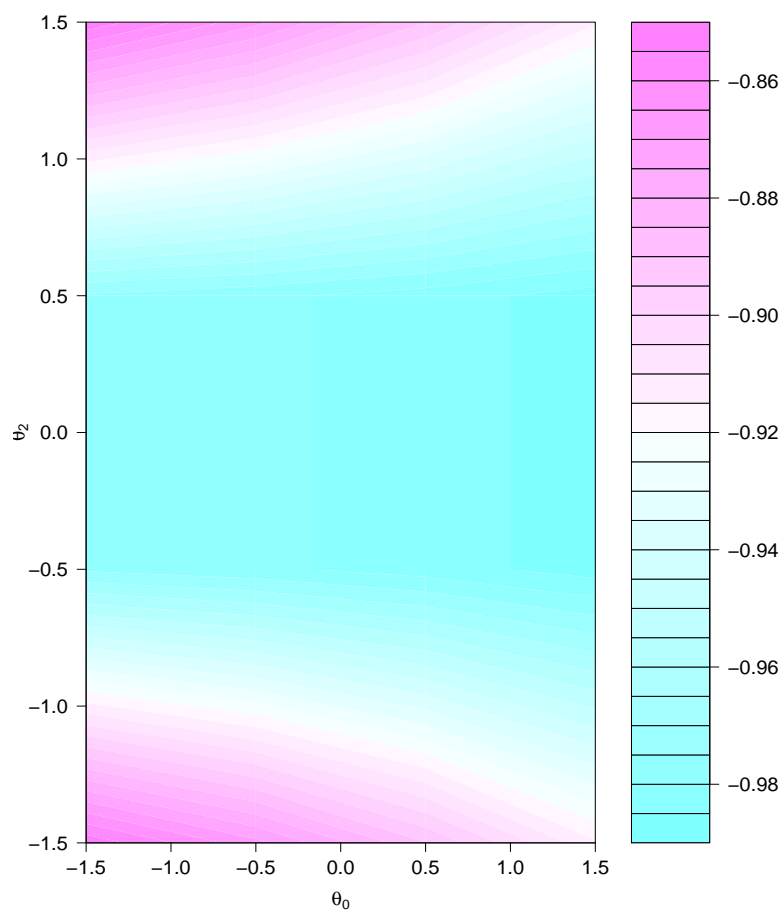
6.4 Discussion and Conclusion

In this chapter we calculated the expected values of terms in equation (6.8). We compared these expected values under different dropout models to the generating values β_3^G and β_4^G. We calculated the bias in the expected parameters in each dropout model. It is clear now that the LME works under MAR, since $\beta_3^* = \beta_3^G$ and $\beta_4^* = \beta_4^G$ or in other words, the asymptotic bias is zero. The maximum likelihood method fails to give a consistent estimate for β_3 or β_4 under MNAR. It is interesting to note one exception that the LME method gives consistent estimates for β_4 under SP dropout model, as happened before for the IPW and CC methods. This may be correct only for the simulation set up used here, but not in general; we have not been able to show theoretically that there is no asymptotic bias.

To conclude the chapter, we summarize and compare all the least false values derived using the LI, IPW, CC and LME methods under SP, MAR and MNAR dropout models. We use the expressions derived in Chapter 3, 4, 5 and the current Chapter 6. Results are in Tables 6.6, 6.7 and 6.8 respectively. We use $\beta = (-2, -2, -1, -1)$, $\theta_0^{sp} = 0$, $\theta_1^{sp} = \sqrt{0.5}$, $\sigma_x^2 = 1$, $\sigma_U = \sigma_{\epsilon_1} = \sigma_{\epsilon_2} = \sqrt{0.5}$. This combination gives the total amount of the dropout at about 50%.

Table 6.6: Comparing the limiting values β_3^* and β_4^* using different methods under SP dropout model

	LI	IPW	CC	LME
β_3^*	-1.00	-0.85	-0.84	-0.92
β_4^*	-1.00	-1.00	-1.00	-1.00

Table 6.7: Comparing the limiting values β_3^* and β_4^* using different methods under MAR dropout model

	LI	IPW	CC	LME
β_3^*	-1.12	-1.00	-0.90	-1.00
β_4^*	-1.05	-1.00	-0.97	-1.00

Table 6.8: Comparing the limiting values β_3^* and β_4^* using different methods under MNAR dropout model

	LI	IPW	CC	LME
β_3^*	-0.89	-0.82	-0.77	-0.83
β_4^*	-0.98	-0.96	-0.95	-0.96

The linear increment model performs well under shared parameter dropout model (refer to Chapter 3). Under MAR, with these parameter settings the estimates are biased down-

wards (more negative), whereas under MNAR they are biased upwards. This is because for LI the estimates at time 2 are based on the differences $Y_2 - Y_1$, for completers. Having lower than expected Y_1 for completers leads to higher than expected differences and downwards bias under MAR, the opposite happens for MNAR.

The IPW method is consistent under MAR, as required but has slightly more bias than LI for β_3 but slightly more for β_4, in our settings. It performs well for β_4 under SP but has positive bias for β_3. The same, under SP, happens for the complete case method CC. This technique is biased under MAR, and badly biased under MNAR. Finally, the LME method performs similarly to IPW in the study, at least in terms of least false values. We have seen previously that it can produce more uncertain estimates however.

Having obtained least false values, in the next chapters we suggest their use in sensitivity analyses. Before doing so, we will investigate a sensitivity procedure for local misspecification as proposed by Copas and Eguchi [2005].

Chapter 7

The effect of local misspecification of the dropout model when using likelihood based methods under the MAR assumption

7.1 Introduction

In the previous chapters, we investigated the consequences of misspecifying the missingness mechanism by deriving the so called least false values, which are the values the parameter estimates converge to when the assumptions may be wrong. The knowledge of these least false values allows us to conduct sensitivity analysis which will be illustrated here for the LME method in this chapter and also for the IPW method in the next chapter. In both chapters we will assume the misspecified dropout model is MAR.

As an alternative, Copas and Eguchi [2005], give a formula to estimate the bias under such misspecification using a likelihood approach. As the LME is a likelihood based method, we can compare the estimates obtained through the Copas and Eguchi method with the LME least false estimates. Also, this idea can be adapted for the IPW estimating equation approach. The procedure will be applied by adding a tilt to the MAR dropout model to provide what Copas and Eguchi call local misspecification.

As we found in Chapter 2, we know that the CC, Obs, and LOCF methods do not work under the MAR dropout model, because they gave biased estimates. So we will concentrate only on the LME and IPW methods for now.

In this chapter, we will elaborate the local model uncertainty as proposed by Copas and Eguchi [2005], and illustrate how to apply it both when model misspecification is present

and when the data is incomplete. Furthermore, we will show that the Copas and Eguchi method gives very similar results to those we found in the previous chapter. Misspecification will be dealt with assuming MAR where actually the truth is MNAR. Beside Copas and Eguchi [2005], many other works developed methods to assess the sensitivity of inference under the MAR assumption; see for example, Luna and Lundin [2014] and Rosenbaum [2010]. Moreover, Lin et al. [2012] extended the Copas and Eguchi method and assumed a doubly misspecified model while we have only single missspecification. There also has been interest in the Copas and Eguchi [2005] method from a Bayes perspective, see for example Zhu et al. [2014], Daniels and Hogan [2008] and Kosuke et al. [2008].

A description of the Copas and Eguchi method is provided in Section 7.2, followed by an example for the two timepoints situation in Section 7.3. A simulation study is described in Section 7.4, the aim being first to make sure that the Copas and Eguchi method is working and then show the coverage of nominal confidence intervals. A sensitivity analysis is conducted to assess how inference can depend on missing data. In Section 7.5, we first apply the methods to the real data example from a clinical trial with two treatments and two measurement times as introduced and analysed by Matthews et al. [2012]. This data was described in Chapter 2. We compare the results obtained by our method with the results found by using the Copas and Eguchi method. Then we apply the methods to the three timepoints real data example. Finally, conclusions are given in Section 7.6.

7.2 Description of Copas and Eguchi Method

We use the notation of Copas and Eguchi [2005], denoting Z for complete data and Y for incomplete data. We have two types of model, the true model and the assumed model. The true model is also called the generating model and it means how the data are actually generated or simulated. On the other hand, the assumed model or what is also known as the fitting model is what we fit to data. The true model for complete data is denoted by $g_Z = g_Z(z; \psi)$ and the corresponding true model for incomplete data is $g_Y = g_Y(y : \psi)$ which can be derived from g_Z. Here ψ is a generic (vector) parameter. The assumed or working model is a parametric model $f_Z = f_Z(z; \psi)$ which gives the distribution of Z, and its marginal density is $f_Y = f_Y(y : \psi)$.

Thus

$$f_Y = \int\limits_{(y)} f_Z \, \mathrm{d}z \tag{7.1}$$

where the notation (y) means integration over all missing values in Z that are consistent with the observed Y.

We will provide a method to approximate the bias in the estimation of the parameters of the misspecified model following Copas and Eguchi [2005]. We will consider MAR as

the working model and MNAR as the true model. Thus the misspecification is caused by assuming MAR but the truth is MNAR.

Suppose we have a random sample of n observations, and the true model is given by g_Z which is defined by equation (16) in Copas and Eguchi [2005] as a tilt model:

$$g_Z = g_Z(z; \psi, \epsilon, u_Z) = f_Z(z; \psi) \exp\{\epsilon u_Z(z; \psi)\}. \tag{7.2}$$

Thus the misspecification is determined by the quantity $\epsilon u_Z(z; \psi)$. In this, ϵ, which is assumed to be small, measures the size of misspecification while $u_Z(z; \psi)$ determines its direction. We assume that $u_Z(z; \psi)$ has zero mean and unit variance under the working model f_Z. The misspecification is local because ϵ is small. Hence, we can infer that g_Z is close to f_Z, and we can write:

$$\frac{g_Z}{f_Z} = \exp\{\epsilon u_Z(z; \psi)\}.$$

Now if we actually use the model $f_Z(z; \psi)$ to fit the data, then the limiting value of the MLE $\hat{\psi}$ as $n \longrightarrow \infty$ is given by equation (18) in Copas and Eguchi [2005] as:

$$
\begin{aligned}
\psi_{g_Z} &= \arg_\psi \left[E_g\{s_Z(z; \psi)\} = 0 \right] \\
&= \psi + \epsilon I_Z^{-1} E_{f_Z} \{u_Z(z; \psi) s_Z(z; \psi)\},
\end{aligned}
$$

where $s_Z(\cdot; \psi) = \partial\{\log(f_Z)\}/\partial\psi$ and $I_Z = E[-\partial^2\{\log(f_Z)\}/\partial\psi\partial\theta^T]$ are the score and information matrix for the model f_Z respectively.

But we will fit to f_Y, the working model for the marginal data. Copas and Eguchi [2005] show that if equation (7.2) is true and ϵ is small then a similar approximation holds for the marginal data Y, i.e.

$$g_Y = g_Y(y; \psi, \epsilon, u_Y) = f_Y(y; \psi) \exp\{\epsilon u_Y(y; \psi)\} \tag{7.3}$$

where again $u_Y(y; \psi)$ has zero mean and unit variance. In this case according to equation (19) in Copas and Eguchi [2005] the limiting value is:

$$\psi_{g_Y} \approx \psi + \epsilon I_Y^{-1} E_f[u_Y s_Y] = \psi + I_Y^{-1} E_f[\epsilon u_Y s_Y] \tag{7.4}$$

where $s_Y(\cdot; \psi) = \partial\{\log(f_Y)\}/\partial\psi$ and $I_Y = E\left[-\partial^2\{\log(f_Y)\}/\partial\psi\partial\psi^T\right]$, are the score and information matrix for the model f_Y respectively. To calculate the bias, $I_Y^{-1} E_f[\epsilon u_Y s_Y]$, we need to find the *tilt* ϵu_Y. In the next section we will determine how to calculate this amount under MAR and MNAR in our setting of two timepoints.

7.3 Copas and Eguchi Method for Two Timepoints Example

The bias consists of, as shown in equation (7.4), the score, information matrix and the tilt. In order to calculate these components we need first to define the likelihood model we will use. Under MAR we can choose either of the following equivalent formulations:

$$L = (f(Y_1, Y_2)P(R = 1|Y_1, Y_2))^R (f(Y_1)P(R = 0|Y_1))^{1-R} \qquad (7.5)$$

$$= (f(Y_2|Y_1)f(Y_1)P(R = 1|Y_1, Y_2))^R (f(Y_1)P(R = 0|Y_1))^{1-R}. \qquad (7.6)$$

The conditional distribution of Y_2 given Y_1 is needed quite a lot in this section. Hence, for simplicity we will use Y_{21} to denote this quantity. Since $f(Y_1, Y_2)$ is bivariate normal in our assumed model we have $Y_{21} \sim N(\mu_{21}, \sigma_{21})$ where $\mu_{21} = \mu_2 + \frac{\sigma_2}{\sigma_1}\rho(Y_1 - \mu_1)$ and $\sigma_{21} = \sigma_2\sqrt{1 - \rho^2}$. Also, we have that the complete data is $Z = (Y_1, Y_2, R)$ and incomplete data is $Y = (Y_1, Y_2^{(R)}, R)$ where

$$Y_2^{(R)} = \begin{cases} Y_2, & R = 1 \\ \\ \text{undefined}, & R = 0. \end{cases} \qquad (7.7)$$

Therefore, at $R = 1$, $Y=Z$, but Y will differ from Z at $R = 0$.

In addition, we define the models f_Z, f_Y, g_Z and g_Y. We assume MAR as the working model or misspecified model. As we know, under MAR we have $P(R = 1|Y_1, Y_2) = P(R = 1|Y_1)$, then from equation (7.5) the working model for complete data by assuming $R = 1$ is:

$$f_Z = f(Y_1, Y_2)P(R = 1|Y_1).$$

Similarly, from equation (7.6) the working model for incomplete data by assuming $R = 0$ is:

$$f_Y = f(Y_1)P(R = 0|Y_1).$$

Under MNAR, if we have complete data then we can always set $R = 1$. Thus, from equation (7.5), the true model for complete data is

$$g_Z = f(Y_1, Y_2)P(R = 1|Y_1, Y_2).$$

The true model for incomplete data on the other hand is the marginal density:

$$g_Y = \int_y g_Z \, \mathrm{d}z$$

$$= \int_{Y_2^R} f(Y_1, Y_2)P(R = 1|Y_1, Y_2) \, \mathrm{d}Y_2^R. \qquad (7.8)$$

Note that the integral is over the missing values Y_2^R. Referring to equation (7.7), we have that the missing values Y_2 are undefined in case that $R = 0$.

This means that in order to use Copas and Eguchi's ideas we need to convert the specific g_Y in equation (7.8) into the general form of equation (7.3). To do this we will redefine our MNAR model in tilt form:

$$P(R = 1|Y_1, Y_2) = \text{expit}\{\theta_0 + \theta_1 Y_1\} \exp\{\epsilon u_Y\}. \tag{7.9}$$

where $\epsilon = \theta_2 \sigma_{21}$ and $u_Y = u_Y(y; \theta) = (Y_2 - \mu_{21})/\sigma_{21}$. For small θ_2 this is a good approximation to the logistic MNAR model used in previous chapters. A more refined approximation is needed in the next chapter.

Calculation of the terms needed for the bias expression (7.4) is now possible and follows directly. Details are omitted.

7.4 Simulation Study

In this section we explore how the parameter estimates are affected when we fit a MAR model to data that are MNAR, and we compare with the values that the Copas and Eguchi method predicts. The simulation set up used here is $\beta = (-2, -2, -1, -1)$, $\theta = (\theta_0, \theta_1) = (-0.5, -0.5)$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1$, $\rho = 0.5$. This gives dropout rate $\approx 40\%$. Also, based on equation (7.9), we use the dropout model:

$$P(R = 1 \mid Y_1, Y_2) = \pi(\theta, \theta_2) = \text{expit}\left(\theta_0 + \theta_1 Y_1 + \theta_2\{Y_2 - \mu_{21}\}\right).$$

where $\mu_{21} = \mu_2 + \rho \sigma_2 (Y_1 - \mu_1)/\sigma_1$.

The first aim is to show that the Copas and Eguchi method works under different simulation parameters. This can be verified by comparing the results obtained by the Copas and Eguchi method with the maximum likelihood parameter estimates. The following plots suggest that the method of Copas and Eguchi works under different parameter sets and dropout rates. Each plot is based on 20 simulations at sample size of 10000. There are three lines. The dotted lines (blue lines) show, for each choice of θ_2 over a grid from -0.2 to 0.2, the mean estimated values of parameters under a MAR assumption when the data are really MNAR. The parameters we consider are: β_3, β_4, σ_2^2 and ρ. We do not give results for β_1, β_2, and σ_1^2 because they are well estimated from the first timepoint observation, which we have assumed to be fully observed. The solid lines (red lines) show mean values derived from equation (7.4), using u_Y and the working score and information. The horizontal lines are at the true values. Figures from 7.1 to 7.4 differ in the parameter sets used in the simulations. Dropout rates vary between figures, with the highest rate in Figure 7.3.

We fit the data under a MAR assumption using maximum likelihood and use this fitted line as a reference. Then, we compare this with the estimates produced using the Copas and Eguchi method. We find that the dashed line matches the solid one in each Figure from 7.1 to 7.4. Clearly, this means that the method of Copas and Eguchi works under MAR. Hence, we can be confident that Copas and Eguchi bias approximation is accurate, at least for small θ_2.

Note that in all Figures from 7.1 to 7.4, at $\theta_2 = 0$ the horizontal lines cross the Copas and Eguchi estimate lines for all parameter estimated. The reason is that at $\theta_2 = 0$, the MAR assumption is valid and thus the estimate will be close to the true value. For example, in Figure 7.1, when looking at the top-left plot, we can see that at $\theta_2 = 0$ the Copas and Eguchi estimator for β_3 equals -1.00 (at the horizontal line) which is the true value of β_3. Moreover, we note that the β_3 and σ_2^2 estimates keep the same pattern in all four figures. In contrast, the β_4 and ρ patterns depend upon the parameter sets. For example, in comparing the β_4 pattern in Figures 7.1 and 7.3, it is clear that in Figure 7.1 the estimates of β_4 decrease as θ_2 increases, whereas in Figure 7.3 the estimates of β_4 increase as θ_2 increases. A similar note can be made for ρ.

Figure 7.1: Default parameter set $\beta = (-2, -2, -1, -1)$, $\theta = (-0.5, -0.5)$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1$, $\rho = 0.5$. This gives dropout rate $\approx 40\%$. Dashed lines (blue lines): Mean of 20 simulations of sample size 10000. Solid lines (red lines): Means of corresponding Copas and Eguchi approximations. Horizontal dotted lines are at the true values.

Figure 7.2: Parameter set $\beta = (-1, -1, 1, 2)$, $\theta = (-0.5, -0.5)$, $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, $\rho = 0.5$. This gives dropout rate $\approx 50\%$. Lines as in Figure 7.1.
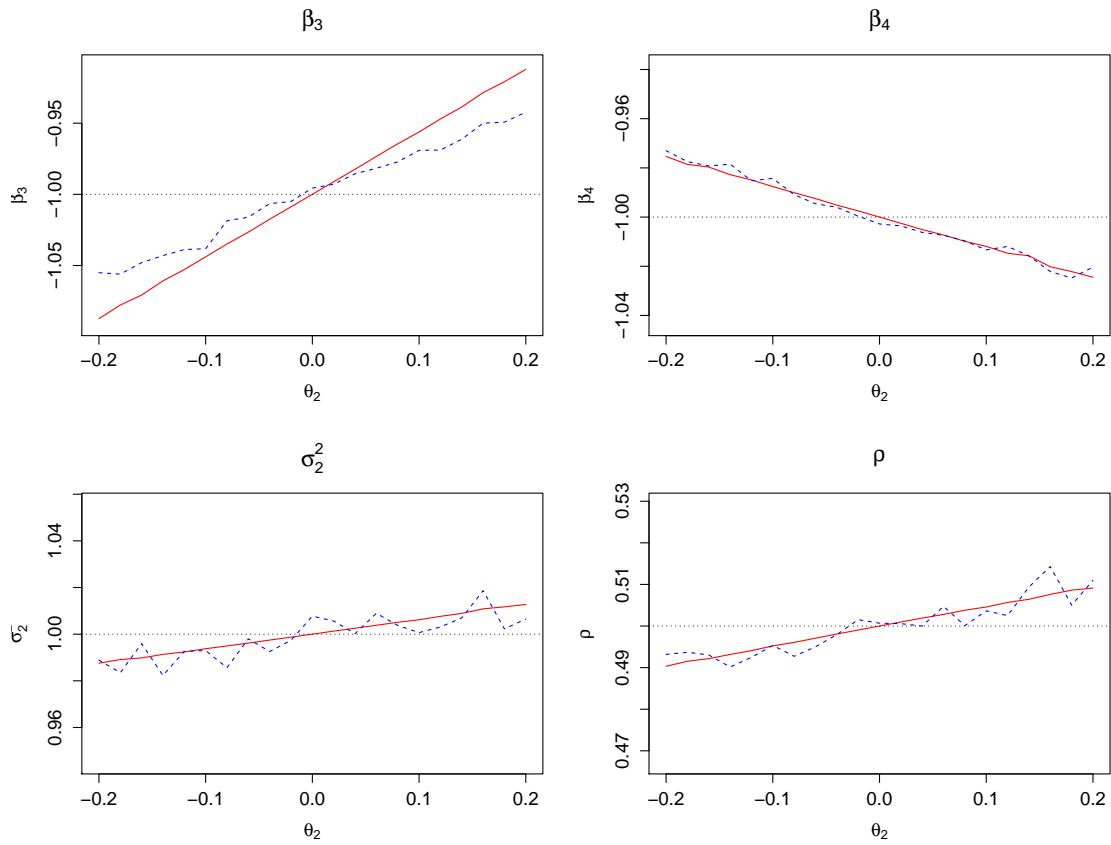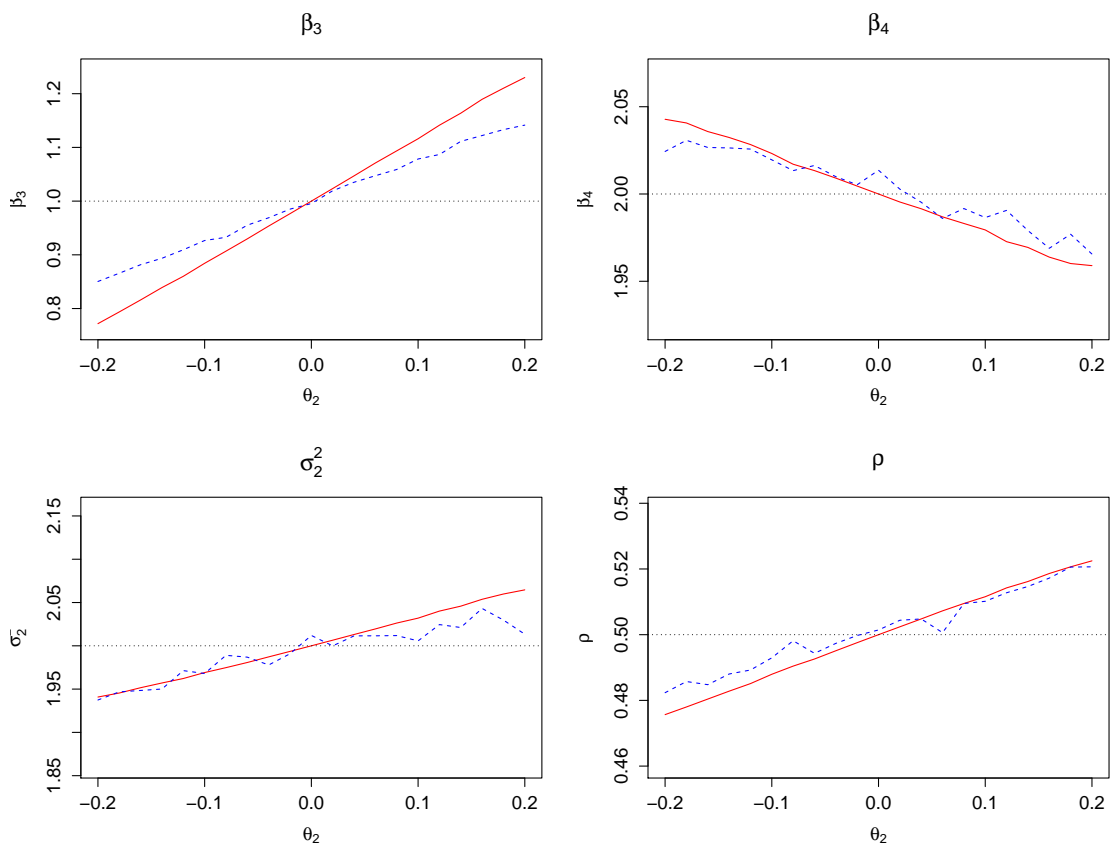
Figure 7.3: Parameter set $\beta = (-2, -2, -1, -1)$, $\theta = (-1, 0.5)$, $\sigma_1^2 = 2$, $\sigma_2^2 = 1$, $\rho = 0$. This gives dropout rate $\approx 83\%$. Lines as in Figure 7.1.
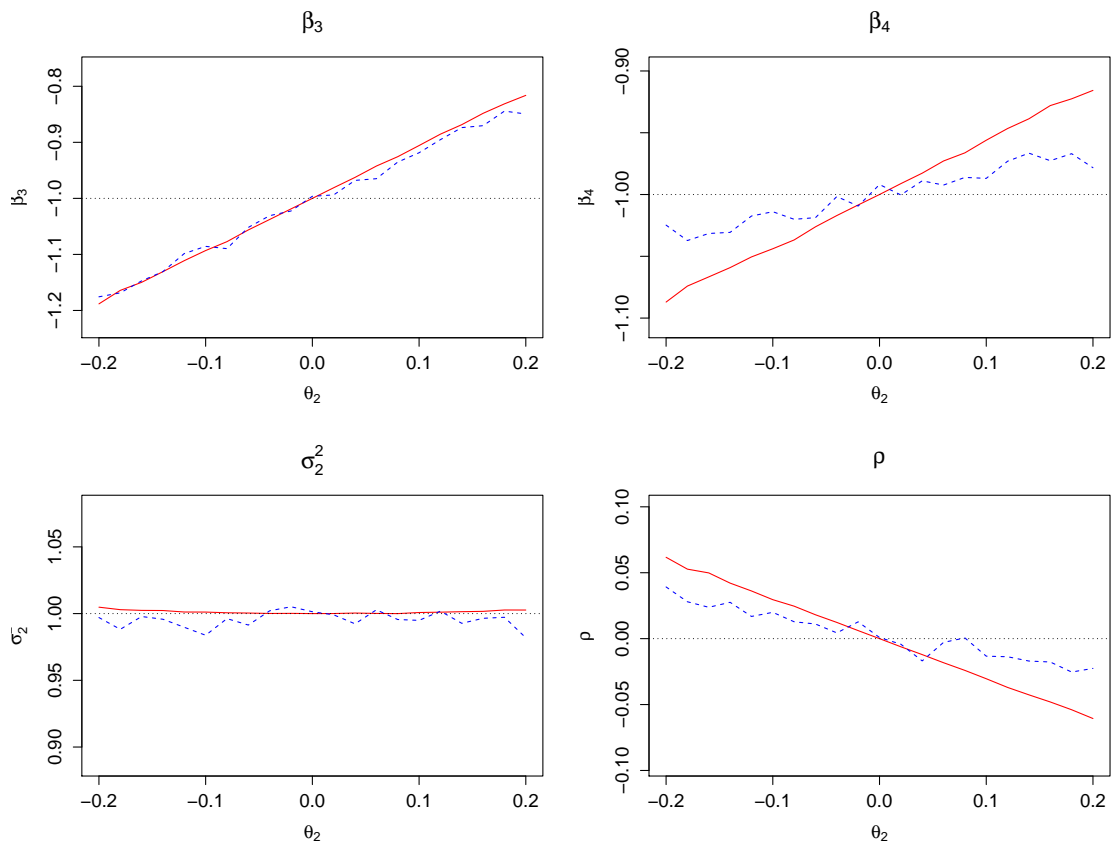
Figure 7.4: Parameter set $\beta = (-2, 0, -2, 0)$, $\theta = (-0.5, -0.5)$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1$, $\rho = 0.75$, and this set leads to dropout rate $\approx 48\%$. Lines as in Figure 7.1.
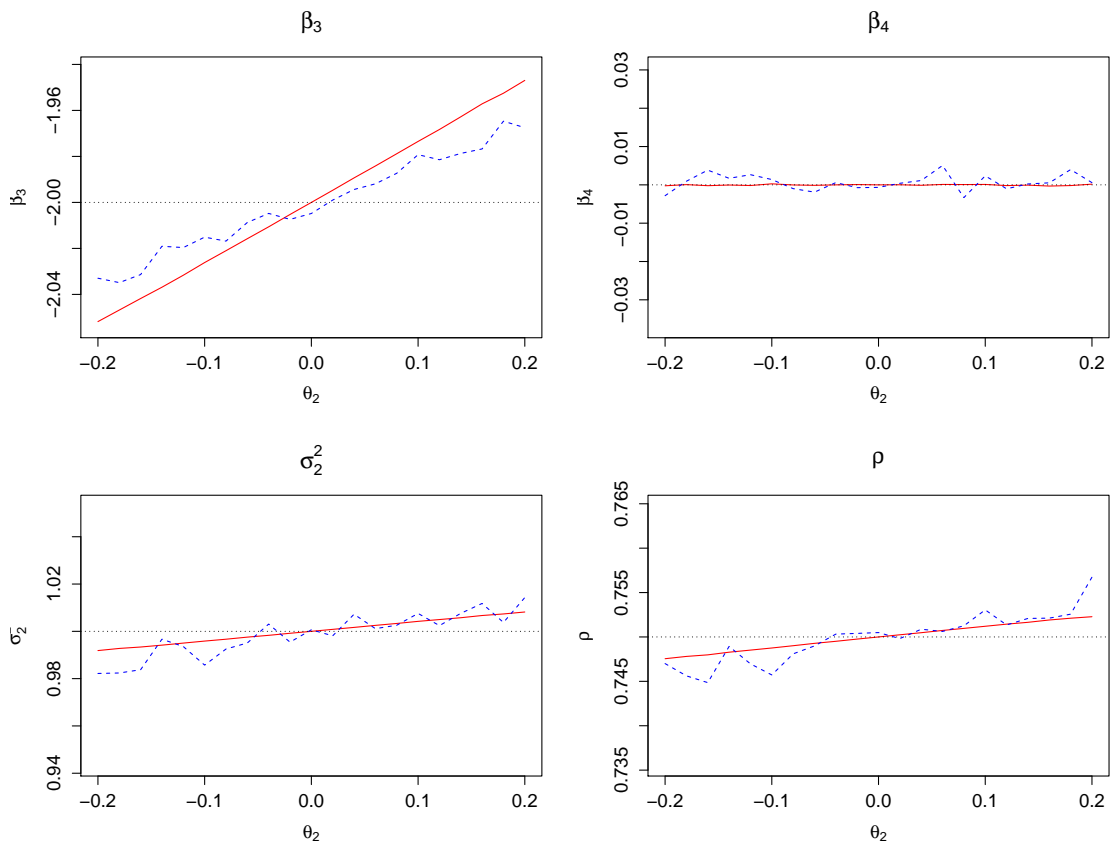
7.4.1 Comparing the Copas and Eguchi method with LME least false results

We want now to compare the LME least false results (from Chapter 6 equation (6.8)) with the Copas and Eguchi [2005] ones. We use sample size 10000 and 10 simulations. The aim is to show the variation in treatment effect estimates as θ_2 varies. We use a grid of θ_2 from -0.2 to 0.2 as previously. We show two examples: Figure 7.5 is produced when we use β=(-2,-2,-1,-1), θ=(-0.5,0) which gives dropout rate around 60% and Figure 7.6 is for β=(-1,-1,-1,-1), θ=(-1,0) which gives a higher dropout percentage of around 83%. In both figures we use $\theta_1 = 0$ because the least false calculations under MNAR was based on $\theta_1 = 0$ to make the integrations easier. However, later, we will improve the calculations to find the MNAR least false at any θ_1. Here the blue lines (dotted lines) are simulation estimates using maximum likelihood, the red lines (solid lines) are Copas and Eguchi estimates, and the light blue lines are the LME least false estimates. These show that the least false, simulations and Copas and Eguchi results all match well. We can therefore use the least false results for bias correction as an alternative to Copas and Eguchi. We focus only on β_3 and β_4 because the LME method is based only on the Y values, also the LME least false we found in the previous chapter are only β_3^* and β_4^*.

Figure 7.5: Comparison 1: $\beta = (-2, -2, -1, -1)$, $\theta = (-0.5, 0)$. The blue lines (dotted lines) are simulation estimates using maximum likelihood, the red lines (solid lines) are Copas and Eguchi estimates, and the light blue lines are the LME least false estimates.
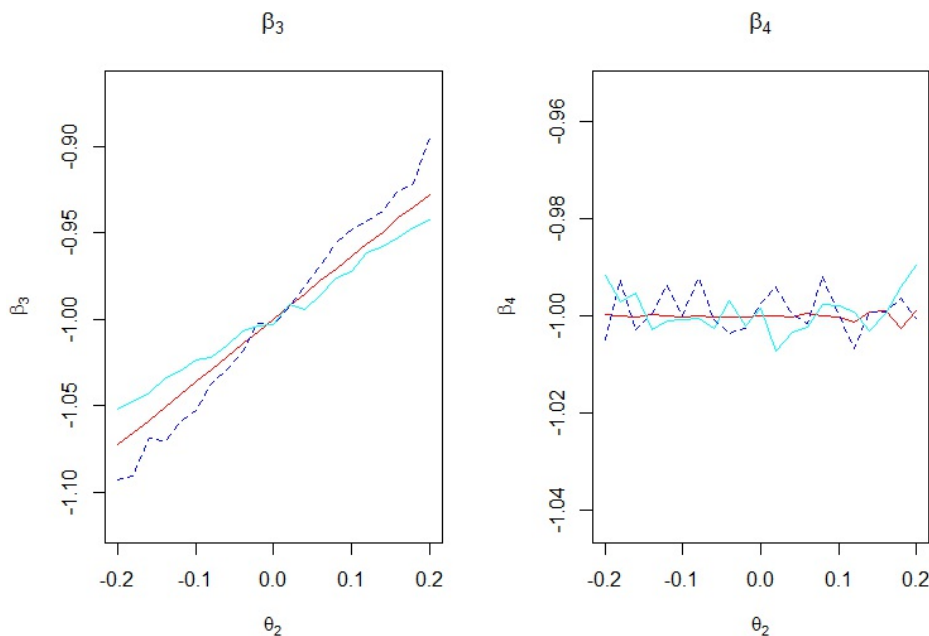
Figure 7.6: Comparison 2: $\beta = (-1, -1, -1, -1)$, $\theta = (-1, 0)$. The lines are as in Figure 7.5.
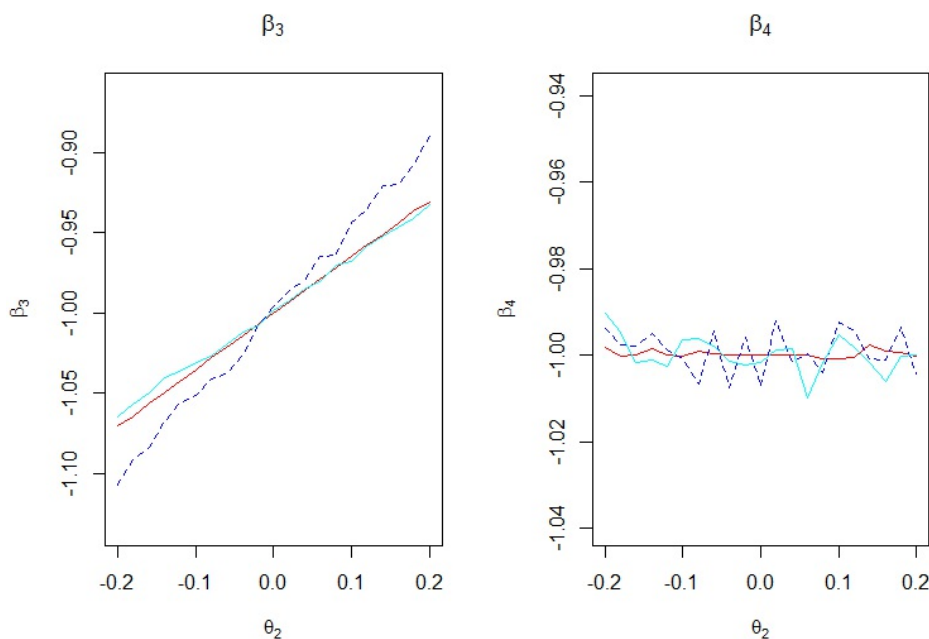
7.4.2 CI coverage for the estimated β_3 and β_4

The Copas and Eguchi and least false values show how estimates are biased by assuming MAR when the data are MNAR. The misspecification parameter is θ_2, with $\theta_2=0$ meaning no misspecification. If the value of θ_2 was known then we could adjust our parameter estimates to take into account the misspecification. This idea will be illustrated in this section.

For a range of true (generating) θ_2 we simulate 1000 samples, each of size 1000. In each case we estimate β_3 and β_4 using maximum likelihood under a MAR assumption. We then adjust the estimates using either the estimated Copas and Eguchi bias or the bias arising through our least false calculations, in both cases taking an *assumed* θ_2. Coverage of the resulting nominal 95% confidence intervals is then recorded. We do not adjust the estimated confidence interval width, just its location.

Tables 7.1 - 7.3 give the results. Here we use θ_2^T for the true θ_2, θ_2^A denotes the assumed value in adjusting the estimates. Also, we use $(\beta_3^{**}, \beta_4^{**})$ for the Copas and Eguchi adjustment method and (β_3^*, β_4^*) for the least false adjustment method.

In Table 7.1 the assumed θ_2 is zero, meaning we make no correction. Results at the correct value of $\theta_2^A =0$ are good. Otherwise, the CI for β_3 goes badly wrong. Note that there is no correction here, so the Copas and Eguchi and least false results should be the same. Small differences are just because of the different calculations that are involved. For example

the least false calculation needs an estimate of σ_x but Copas and Eguchi does not. We note that the CI coverage for β_4 is not too much affected at any true θ_2 in the range (-0.1,+0.1) that we looked at. For example, at θ_2^A=-0.1, the CI coverage for β_4 is about 95%, whereas there is undercoverage for β_3 when θ_2^T deviates from zero. For example at θ_2^T=-0.1, the CI coverage for β_3 is about 85%. This indicates that β_4 is less sensitive for the misspecification than β_3 in this scenario.

In Table 7.2 we take the assumed value of θ_2=-0.1, meaning we think dropout is associated with high Y_2. Things work well if our assumed value is close to the true one, -0.1. Note that in contrast to the Table 7.1, there is correction here, so the Copas and Eguchi and least false results will not be the same, for example at θ_2^T=+0.1, the CI coverage for β_3^{**} is about 62.7%, but the CI coverage for β_3^* is about 57.8%. However, both estimates β_3^{**} and β_3^* have undercoverage as θ_2^T goes further from the assumed value -0.1.

We provide one more example, when the assumed θ_2 is positive. From Table 7.3, we have the same conclusion as before. For example, when the assumed value is close to the true value, θ_2^T=+0.05, there is good coverage for CI for the estimates. In contrast, at θ_2^T=-0.10, the CI coverage for β_3^{**} is about 75.1% and the CI coverage for β_3^* is about 72.5%. Again, β_4 seems very robust.

Table 7.1: CI coverage in percent for the estimated β_3 and β_4 at assumed $\theta_2{=}0$. We use θ_2^T for the true θ_2, θ_2^A for the assumed value in adjusting the estimates, $(\beta_3^{**}, \beta_4^{**})$ for the Copas and Eguchi adjustment method and (β_3^*, β_4^*) for the least false adjustment method. Results based on 1000 samples of size 1000.

θ_2^T	θ_2^A	β_3^{**}	β_4^{**}	β_3^*	β_4^*
-0.10	0.00	84.80	95.30	84.90	95.30
-0.09	0.00	85.90	96.70	85.80	96.70
-0.08	0.00	86.80	95.30	86.60	95.20
-0.07	0.00	89.70	94.40	89.70	94.40
-0.06	0.00	92.00	94.70	91.90	94.70
-0.05	0.00	92.00	96.00	91.90	96.00
-0.04	0.00	93.70	95.30	93.60	95.30
-0.03	0.00	95.00	95.00	95.10	94.90
-0.02	0.00	94.60	94.40	94.60	94.60
-0.01	0.00	95.90	96.10	95.70	96.00
0.00	0.00	94.70	95.10	94.70	95.10
0.01	0.00	94.30	95.20	94.40	95.20
0.02	0.00	95.00	95.20	95.00	95.20
0.03	0.00	95.20	94.20	95.00	94.20
0.04	0.00	93.90	94.70	93.90	94.70
0.05	0.00	94.10	95.00	94.00	94.90
0.06	0.00	91.70	94.70	91.70	94.70
0.07	0.00	90.30	95.40	90.20	95.40
0.08	0.00	88.20	94.90	88.20	94.90
0.09	0.00	88.00	95.00	87.80	95.00
0.10	0.00	83.40	95.10	83.60	95.00

Table 7.2: CI coverage for the estimated β_3 and β_4 in percent at assumed θ_2=-0.1. We use θ_2^T for the true θ_2, θ_2^A for the assumed value in adjusting the estimates, $(\beta_3^{**}, \beta_4^{**})$ for the Copas and Eguchi adjustment method and (β_3^*, β_4^*) for the least false adjustment method.

θ_2^T	θ_2^A	β_3^{**}	β_4^{**}	β_3^*	β_4^*
-0.10	-0.10	95.30	95.40	95.70	95.10
-0.09	-0.10	94.80	95.10	95.50	94.80
-0.08	-0.10	94.10	96.10	94.10	96.00
-0.07	-0.10	95.50	96.70	95.60	95.80
-0.06	-0.10	95.80	95.20	94.90	95.60
-0.05	-0.10	93.40	94.40	92.00	94.50
-0.04	-0.10	93.70	95.10	92.40	95.20
-0.03	-0.10	92.50	94.90	92.00	95.10
-0.02	-0.10	92.40	94.70	90.50	94.80
-0.01	-0.10	91.20	95.10	89.60	95.30
0.00	-0.10	89.30	95.30	87.40	95.40
0.01	-0.10	89.30	95.80	86.40	96.10
0.02	-0.10	86.80	94.50	83.80	94.60
0.03	-0.10	83.70	93.80	81.20	93.80
0.04	-0.10	80.30	95.50	76.70	95.40
0.05	-0.10	76.00	94.30	71.70	94.50
0.06	-0.10	74.40	95.10	70.50	95.00
0.07	-0.10	69.90	95.70	65.60	95.20
0.08	-0.10	65.20	95.10	61.70	95.00
0.09	-0.10	62.20	96.00	58.30	95.80
0.10	-0.10	62.70	93.90	57.80	94.20

Table 7.3: CI coverage for the estimated β_3 and β_4 in percent at assumed θ_2=0.05. We use θ_2^T for the true θ_2, θ_2^A for the assumed value in adjusting the estimates, $(\beta_3^{**}, \beta_4^{**})$ for the Copas and Eguchi adjustment method and (β_3^*, β_4^*) for the least false adjustment method.

θ_2^T	θ_2^A	β_3^{**}	β_4^{**}	β_3^*	β_4^*
-0.10	0.05	75.10	97.20	72.50	97.30
-0.09	0.05	75.50	95.20	73.70	95.40
-0.08	0.05	79.10	95.10	76.50	95.50
-0.07	0.05	82.10	95.10	79.90	95.10
-0.06	0.05	84.10	96.40	81.70	96.20
-0.05	0.05	86.40	96.40	85.00	96.70
-0.04	0.05	88.70	95.20	87.20	95.30
-0.03	0.05	91.10	94.90	89.90	95.00
-0.02	0.05	91.50	95.20	90.50	95.30
-0.01	0.05	92.60	96.20	91.50	96.00
0.00	0.05	94.30	95.40	93.40	95.50
0.01	0.05	95.30	95.70	94.70	95.60
0.02	0.05	94.40	96.00	94.00	96.00
0.03	0.05	94.80	95.70	94.40	95.40
0.04	0.05	96.00	95.60	95.90	95.20
0.05	0.05	95.60	94.90	95.60	94.60
0.06	0.05	95.10	94.40	95.40	94.10
0.07	0.05	93.60	95.30	94.00	95.00
0.08	0.05	93.80	95.00	94.40	95.10
0.09	0.05	92.10	94.60	92.50	94.80
0.10	0.05	91.70	95.10	92.80	95.20

7.4.3   Sensitivity analysis

Of course, in practice θ_2 is not known. For any given data set, a sensible sensitivity procedure would mean plotting bias-corrected estimates and confidence intervals for a range of assumed θ_2 values. Here we use a grid of assumed θ_2 from -0.2 to 0.2. We will show that for each limiting value calculated by the Copas and Eguchi method, the simulated values are within noise of the theoretical values for large sample sizes (n=100000). We estimate the noise from the simulations; that is we get a confidence interval from the simulations and reassurance that the population values are within these. We did this for a correct MAR model and after that under true MNAR but MAR is assumed.

Figure 7.7 illustrates when MAR is the correct model ($\theta_2 = 0$) and the unadjusted confidence intervals (red lines) include the true parameter values (β_3=-1 and β_4=-1), as in this case so do the adjusted ones (blue lines). The horizontal lines are at the true values. We note that β_3^{**} decreases as θ_2 increases whereas β_4^{**} increases as θ_2 increases. This is the opposite to Figure 7.1, which has the same parameter values. The reason is that we are adjusting for the assumed parameter. For example, at $\theta_2 = -0.2$, Figure 7.1 shows that we expect our MAR estimate to be too low. Hence we adjust for it. But because the true θ_2 is zero in Figure 7.7, we mistakenly over-estimate β_3 at θ_2^A=-0.2. The argument also applies to β_4. Note that β_4 has a wider CI than β_3.

Figure 7.8 has the true θ_2=0.1 so we have fitted MAR to data that are really MNAR. The lines cross at θ_2=0 because we are fitting the same MAR model. The important point is that we get better estimates of the true β's at the correct θ_2. Also, as mentioned in Figure 7.7, β_4 has wider CI than β_3.

We note that, both under MAR and MNAR, β_3 and β_4 have opposite trends; β_3 decreases as θ_2 increases whereas β_4 increases as θ_2 increases.

Figure 7.7: CI under MAR: $\beta = (-2, -2, -1, -1)$, $\theta = (-0.5, -0.5)$, $\theta_2^T = 0$. The blue lines are the adjusted estimates, red lines are the unadjusted estimates. The horizontal dotted lines are at the true values.
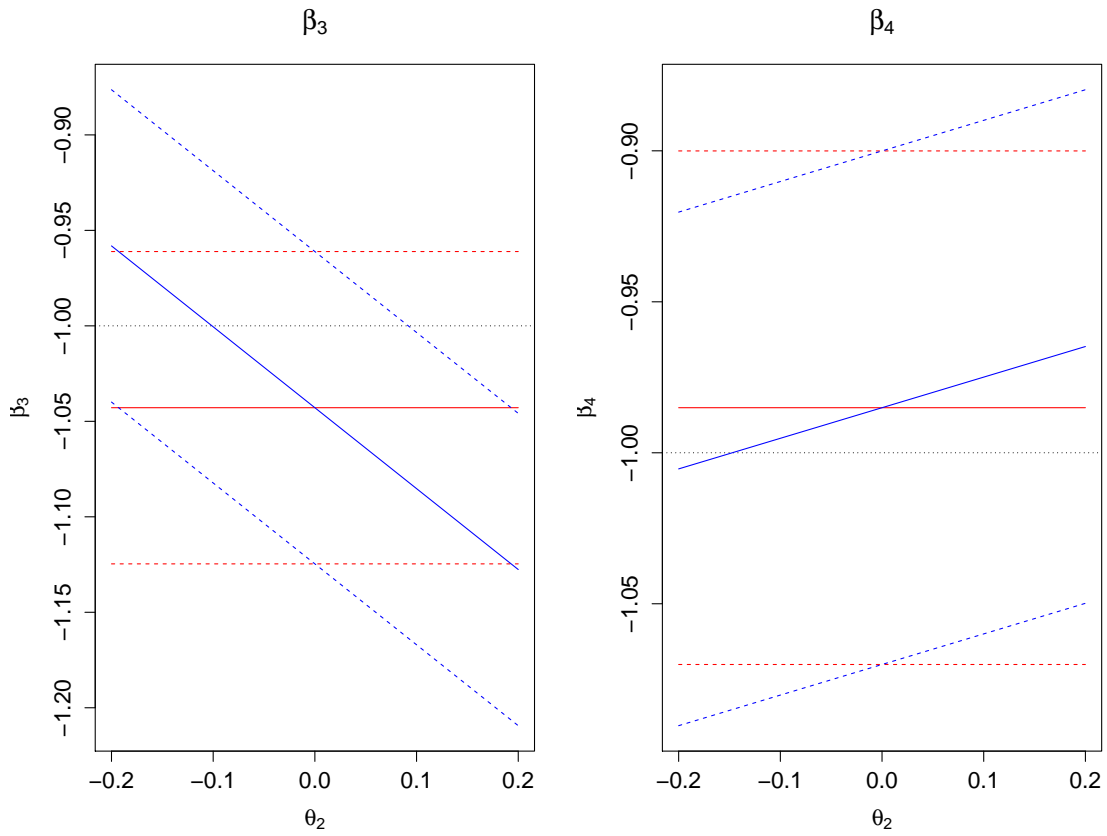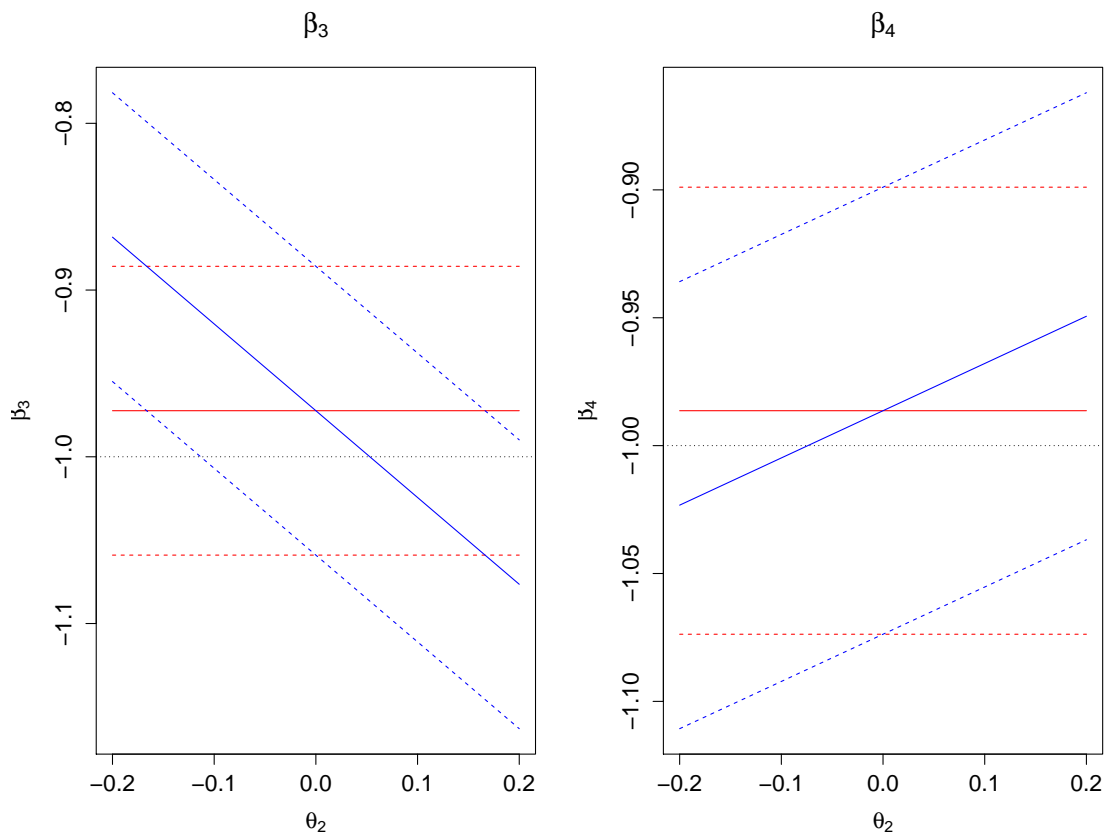
Figure 7.8: CI under MNAR: $\beta = (-2, -2, -1, -1)$, $\theta = (-0.5, -0.5)$, θ_2^T=0.1. The blue lines are the adjusted estimates, red lines are the unadjusted estimates. The horizontal lines are at the true values.

7.5 Application

### 7.5.1 Sensitivity analysis: Two timepoints example

In this section we use the two timepoints real data example, which is the same data as used in Chapter 2. This data is drawn from a clinical trial with two treatments and two measurement times. The covariates are only treatment type and time. The parameter vector is $(\beta_1, \beta_2, \beta_3, \beta_4)$, ignoring any time interaction. There are 422 subjects, assigned to either Treatment A or B. Treatment A is associated with treatment effect $x=1$ and treatment B is when $x=0$. Then, at time 2, the mean of the group receiving Treatment B is β_3 and the mean of the group receiving Treatment A is $\beta_3+\beta_4$. At time 1, all subjects provided a response, but 24.4% dropped at time 2. There are 212 subjects receiving Treatment A, but only 126 provided a response by time 2 and the other 86 dropped out. Hence the missingness percentage is about 40%. The dropout reason is not known. For Treatment B, there are 210 subjects, of which 193 subjects continued to time 2 and hence there are 17 that did not and this gave around 8% missingness.

A sensitivity analysis approach (over a grid of θ_2) using the Copas and Eguchi and LME methods is shown in Figure 7.9. The blue lines use Copas and Eguchi method and the red lines use the least false method. The idea is to adjust the estimate to compensate for bias from a misspecified MAR fit. So, for example, if we know the least false value under MAR underestimates a parameter, we add the difference to our estimate to back-calculate. Dashes are the CIs, based on the MAR standard errors. The first plot shows confidence intervals for the treatment B mean as the assumed value of θ_2 changes. The horizontal line is the estimate under MAR. The second plot shows the confidence intervals for the mean of treatment A. The third plot is the difference in means between Treatment A and B, which yields the treatment effect means i.e. β_4 means. In the first plot, the horizontal line is at -0.74 which is the same value for the LME estimate for β_3 in Chapter 2 Table 2.17. Again, the LME estimate for β_4 is about -0.40 in both Table 2.17 in Chapter 2 and here in Figure 7.9. Also, note that $\beta_3+\beta_4$ equals -1.15. This supports our finding here, and make us more confident about the results.
The first thing to note is how close the least false and Copas and Eguchi estimates are. There is almost no difference over this range of θ_2. We take θ_2 from -1.5 to +1.5. The value of $\hat{\theta}_1$ under MAR is -1.66, meaning the range of θ_2 allows Y_2 to have the same order of effect as Y_1. Clearly at large values of θ_2 we might worry that the misspecification is not local, which is the assumption of Copas and Eguchi. However, the least false results apply to any misspecification, not necessarily local, and the fact that Copas and Eguchi is so close to least false suggests that it can work well even under quite large misspecification.

When θ_2 is negative the estimates get adjusted upwards, the opposite for θ_2 positive. This makes sense: At negative θ_2 large Y_2 values have low probability of staying in the trial. Hence the observed means are lower than they would be in the hypothetical no-dropout

Figure 7.9: Two timepoints example: 95% CI for β_3 and β_4. The blue lines use Copas and Eguchi method, the red lines use least false method and the horizontal line is at the MAR estimate.
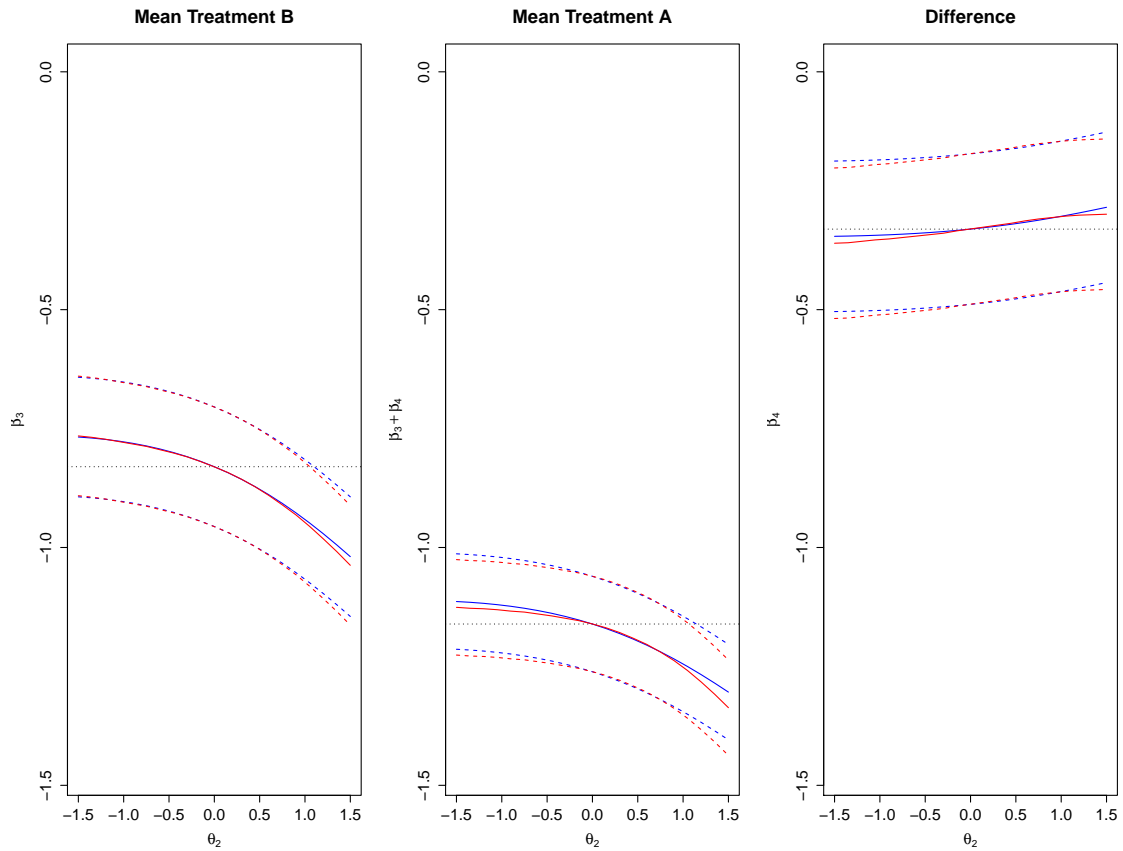
situation, so we adjust upwards.

The estimates seem to be affected more at positive θ_2 than negative. At the very largest θ_2 shown, there would be a significant change in the value of the estimated true mean. However, there is very little effect of misspecification on the difference between means (third sub plot), as the adjustments essentially cancel.

7.5.2 Sensitivity analysis: Three timepoints example

In this section we use the three timepoints schizophrenia data as described in Chapter 2. This data is from a clinical trial with three treatments and three measurement times. The covariates are only treatment type and time. There are three treatment groups: group 1 has 85 subjects, group 2 has 88, group 3 has 345. We will have a separate parameter for the mean of each of the 9 treatment and time combinations. The parameter vector is $(\beta_1, \beta_2, \beta_3)$ at time 1, $(\beta_4, \beta_5, \beta_6)$ at time 2, and $(\beta_7, \beta_8, \beta_9)$ for groups 1, 2 and 3 respectively at each timepoint.

Of the 518 patients, 249 dropped out and the missingness percentages are around 24% and 48% at the second and third time respectively. Examining missingness according to the groups shows that treatment group 2 has the highest dropout percentage while treatment groups 1 and 3 have lower dropout percentage.

We have two dropout models both assumed MAR. The MAR dropout model is (1.6) as defined in the Introduction.

Figure 7.10: 95% CI for the means for groups of schizophrenia data at time 2 and time 3 using Copas and Eguchi and LME methods. The blue lines use Copas and Eguchi method, the red lines use least false method and the horizontal line is at the MAR estimate.
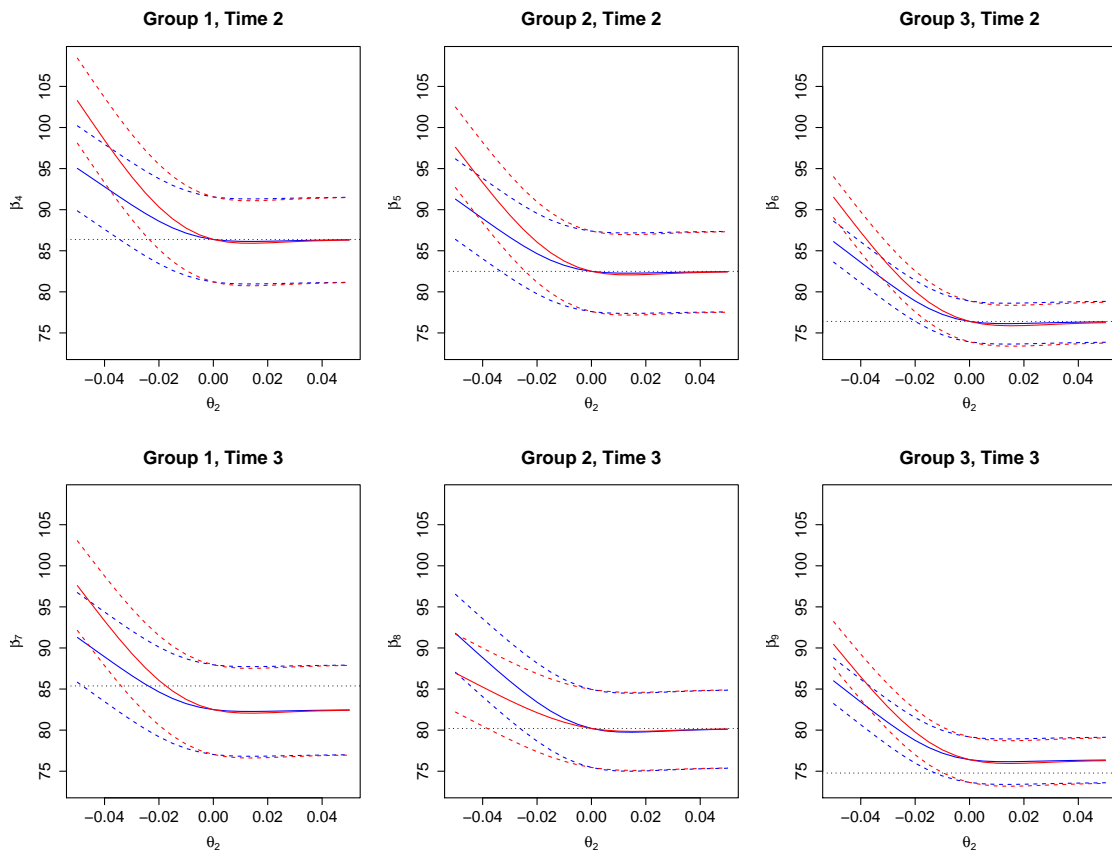
Figure 7.10 shows estimates of the group means (with CIs) at each of time 2 and time 3. The dotted lines are at the MAR estimates. The CIs are wider for groups 1 and 2 than group 3, because group 3 is much larger. The value of θ_1 in the data is around -0.01 (time 2) and -0.03 (time 3). Hence the range of θ_2 is chosen to be -0.05 to +0.05, to be of the same order of magnitude. The Copas and Eguchi method is for local misspecification which implies θ_2 near zero. Note that at positive θ_2 there is no adjustment because there is no bias. The reason is that for the estimated parameters at positive θ_2 we expect (in theory) very little dropout, so there is no need to adjust. Of course in practice there is dropout, which means positive θ_2 is not realistic (in association with the current θ_0 and θ_1). For negative θ_2 all estimates are adjusted upwards, for the same reasons as for the two time example. the least false method seems to lead to larger adjustments than Copas and Eguchi. As mentioned above, Copas and Eguchi is local, meaning if there is a difference for the more extreme θ_2 then probably the least false version should be preferred. The results are slightly different from Chapter 2 (Table 2.19) because we do not estimate at time 1, because at time 1 there is no dropout, so we use the full data and thus no misspecification has occurred. The magnitudes of the adjustment are comparable to those presented (for a joint modelling strategy) by Henderson et al. [2000] in their analysis of an extended version of these data.

Figure 7.11: 95% CI for the differences in means for groups of schizophrenia data at time 2 and time 3 using Copas and Eguchi and LME methods. The blue lines use Copas and Eguchi method, the red lines use least false method and the horizontal line is at the MAR estimate.
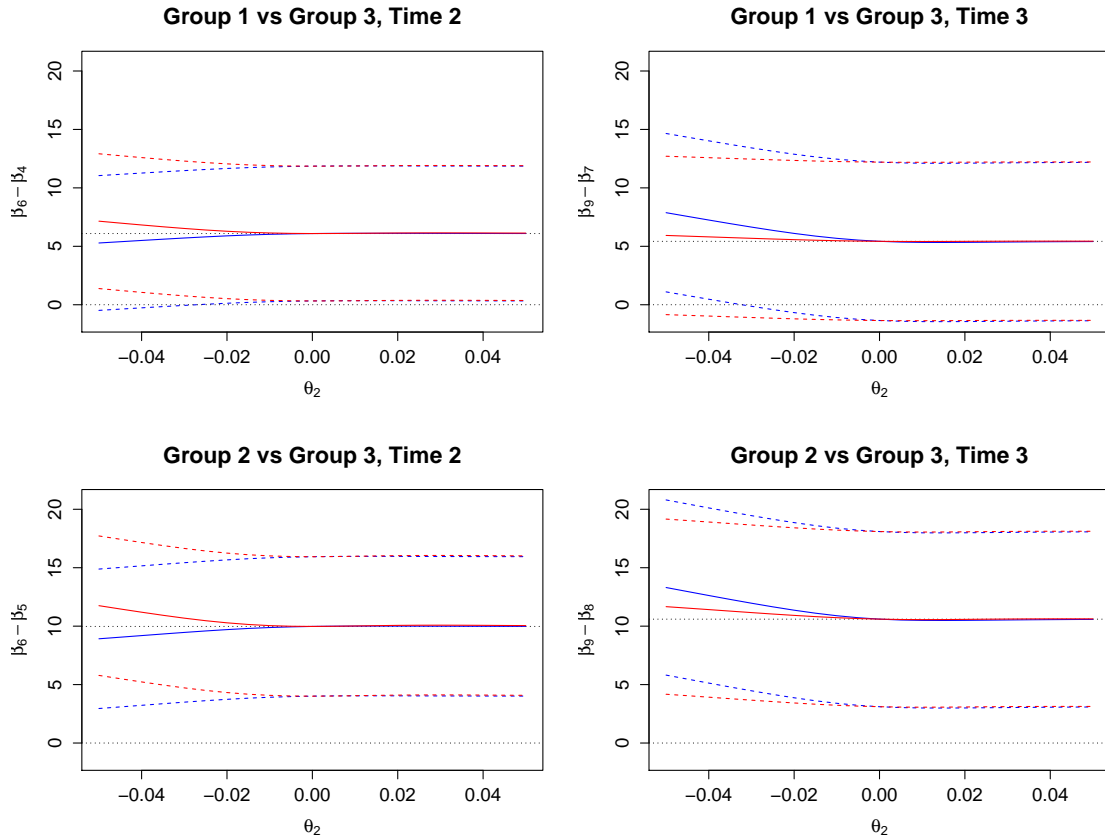

Figure 7.11 shows estimates of the treatment effects (group 1 vs group 3 and group 2 versus group 3) at each of time 2 and time 3. There are now dotted lines at zero as well as the MAR estimate. It is interesting to note that although the mean estimates in Figure 7.10 are severely adjusted, the differences in means have very little adjustment: the two effects cancel. Results confirm that there is a significant difference between group 2 and group 3 at both times, and this is maintained even under MNAR. The difference between group 1 and group 3 is borderline significant under MNAR with θ_2 around -0.04, which would be a very strong effect. High values in these data mean more ill patients. The positive differences mean group 3 (the new treatment) works best. These are the same conclusions as Henderson et al. [2000] obtained. We have obtained them without having to fit a complex MNAR model.

7.6 Conclusion

In this chapter we derived and explored the Copas and Eguchi [2005] approximation for the bias raised by the misspecification of the working model. We compared these results with the LME least false values as obtained in Chapter 6. Moreover, we explained how to use a sensitivity analysis to see how the methods work under a range of θ_2. Finally, we illustrated results using two real examples.

We found that Copas and Eguchi [2005] method and LME least false match very well. Both gave very close results over the grid of θ_2 we considered. This suggests that our least false method can provide a credible alternative to Copas and Eguchi in sensitivity analysis. In fact it might be preferred since there is no assumption of local misspecification.

Chapter 8

The effect of local misspecification on dropout probabilities and IPW estimation

8.1 Introduction

In the previous chapter, we investigated the effect of a tilt or misspecification of the dropout model when using a likelihood based method under the MAR assumption. Specifically, we studied and compared the Copas and Eguchi [2005] bias estimate results with the least false values derived from the LME method. In this chapter, we will take a similar approach, and again we will assume that MAR is the working model while the truth is MNAR. We refer to Ho et al. [2012] and adapt their work to our set up. The objectives here are first to see what the dropout parameters (θ_0 and θ_1) will converge to when using the local misspecification method, then, to estimate the regression parameters β using the IPW method. We develop the Copas and Eguchi [2005] approximation for the non-likelihood estimating equation approach. Furthermore, we show how sensitive the original IPW method is to misspecified models.

This chapter contains two main parts: In the first part we concentrate on the estimation of the dropout parameters of the logistic regression model i.e. θ_0 and θ_1. The second part focuses on using the results obtained in the first part to estimate the regression parameters β when using the IPW estimating equation approach. A full description of the method is shown in Section 8.2. A simulation study for the first part is described in Section 8.3, the aim being first to make sure that the local misspecification method is working and then show the coverage of confidence intervals. A sensitivity analysis is conducted to assess how inference can depend on missing data. We perform the sensitivity analysis by displaying the estimators θ_0 and θ_1 for various choices of θ_2. We apply our approach to the analysis of the real data examples in Section 8.4. Then, in Section 8.5 we do the second part i.e. investigate the IPW method. Again, we apply the methods to both real data examples in

Section 8.6. Finally, we summarize the main ideas discussed in this chapter and highlight the findings in Section 8.7.

8.2 Logistic Regression Method to Estimate the Dropout Probabilities θ_0^* and θ_1^*

In the previous chapter, as mentioned, we concentrated only on the βs. We now consider what happens to the estimated dropout probabilities when the working model is MAR but the true model is MNAR. In the previous chapter, we used the dropout model:

$$P(R = 1 \mid Y_1, Y_2) = \pi(\theta, \theta_2) = \text{expit}\left(\theta_0 + \theta_1 Y_1 + \theta_2 \{Y_2 - \mu_{21}\}\right).$$

where $\mu_{21} = \mu_2 + \rho\,\sigma_2(Y_1 - \mu_1)/\sigma_1$. However, here we revert to the original dropout model. Let $\theta = (\theta_0, \theta_1)$ and

$$P(R = 1 \mid Y_1, Y_2) = \pi(\theta, \theta_2) = \text{expit}\left(\theta_0 + \theta_1 Y_1 + \theta_2 Y_2\right).$$

For small $\theta_2 Y_2$ we can expand[1] $\theta_2 Y_2$ about 0 and get

$$\pi(\theta, \theta_2) \simeq \pi(\theta, 0) + \theta_2 Y_2 \,\pi(\theta, 0)\,(1 - \pi(\theta, 0)). \tag{8.1}$$

We will be fitting the MAR model to MNAR data using maximum likelihood. Ho et al. [2012] explained that the parameter estimates $(\hat{\theta})$ will converge to the values (θ^*) that minimise the expected Kullback-Liebler divergence, and in our case will solve

$$U(\theta^*, \theta, \theta_2) = 0,$$

where $U(\theta^*, \theta, \theta_2)$ is a 2×1 vector

$$\begin{pmatrix} E\left[\pi(\theta, \theta_2) - \pi(\theta^*, 0)\right] \\ \\ E\left[Y_1\{\pi(\theta, \theta_2) - \pi(\theta^*, 0)\}\right] \end{pmatrix}. \tag{8.2}$$

Expanding in θ^* about θ

$$0 = U(\theta^*, \theta, \theta_2) \simeq U(\theta, \theta, \theta_2) + \frac{\partial U}{\partial \theta^*}(\theta^* - \theta). \tag{8.3}$$

Here $\partial U/\partial \theta^*$, evaluated at $\theta^* = \theta$, is the 2×2 matrix

$$\begin{pmatrix} -E\left[\pi(\theta, 0)\,(1 - \pi(\theta, 0))\right] & -E\left[Y_1\,\pi(\theta, 0)\,(1 - \pi(\theta, 0))\right] \\ \\ -E\left[Y_1\,\pi(\theta, 0)\,(1 - \pi(\theta, 0))\right] & -E\left[Y_1^2\,\pi(\theta, 0)\,(1 - \pi(\theta, 0))\right] \end{pmatrix}. \tag{8.4}$$

---

[1]Expanding $f(x)$ about a means $f(x) \approx f(a) + (x - a)\,f'(a)$.

Using (8.1)

$$
U(\theta, \theta, \theta_2) \simeq \begin{pmatrix} E\left[\pi(\theta, 0) - \pi(\theta, 0)\right] \\ \\ E\left[Y_1\{\pi(\theta, 0) - \pi(\theta, 0)\}\right] \end{pmatrix} + \begin{pmatrix} E\left[\theta_2 Y_2 \pi(\theta, 0)\left(1 - \pi(\theta, 0)\right)\right] \\ \\ E\left[\theta_2 Y_2 Y_1 \pi(\theta, 0)\left(1 - \pi(\theta, 0)\right)\right] \end{pmatrix}
$$
$$
= \quad 0 + \psi(\theta, \theta_2) \qquad \text{(say)}. \tag{8.5}
$$

Hence using the result of equation (8.5), equation (8.3) yields:

$$
\theta^* \simeq \theta - \left(\frac{\partial U}{\partial \theta^*}\right)^{-1} \psi(\theta, \theta_2). \tag{8.6}
$$

In practice we replace expectations by sample means, and replace the unknown θ by the estimate $\hat{\theta}$. Note that the approach here, based on small θ_2 and expansion of the estimating equations, is not the same as the previous least false method, which does not assume small θ_2 and *solves* the estimating equations.

8.3 Simulation Study for the Dropout Probabilities θ

The objective here is to compare the estimates obtained by using the local misspecification method with logistic regression least false results for a variety of parameter values. We use the same simulation set up as used in the previous chapter i.e. $\beta = (-2, -2, -1, -1)$, $\theta = (\theta_0, \theta_1) = (-0.5, -0.5)$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1$, $\rho = 0.5$. This gives dropout rate $\approx 40\%$. The dropout model we use here is, as mentioned before,

$$
P(R = 1 \mid Y_1, Y_2) = \pi(\theta, \theta_2) = \operatorname{expit}\left(\theta_0 + \theta_1 Y_1 + \theta_2 Y_2\right).
$$

Results show that both methods work very well, Figure 8.1 and 8.2 illustrate. The left plot refers to θ_0 and the right to θ_1. In each plot the blue line shows the average value of the estimated θ, over 20 simulations, each with sample size $n=10000$, as we vary the true (but ignored[2]) θ_2 from -0.2 to +0.2. The estimate of θ_0 is very stable but the estimate of θ_1 can be highly affected by ignoring θ_2. The red lines in each plot give the expected estimates derived using the (new) Copas and Eguchi method i.e. the local misspecification estimation method, which is expressed in equation (8.6). The green lines give the least false estimates (from Chapter 4 results). Sometimes least false and local misspecification results are so close that the green line is not visible. The horizontal line is at the true value.

Figures 8.1 and 8.2 differ in the true values of θ_0 and θ_1. But they both show that the local misspecification method works, because the estimates derived through equation (8.6) match the other method estimates over the values of θ_2. In Figure 8.1 the horizontal line

---

[2]Because we are assuming MAR, the dropout model does not depend on θ_2

and the estimates lines cross at $\theta_2 = 0$, in each plot. This is because, in theory, at $\theta_2 = 0$ the MAR assumption is valid and so the the estimates will be unbiased which means the estimates will give the true value. For example, in the right plot, at $\theta_2 = 0$ the horizontal line at the true value (at $\theta_1 = -0.5$) crosses the estimates lines, as expected.

Figure 8.1: Parameters $\theta = (-0.5, -0.5)$. The blue line shows the average value of the estimated θ, the red lines give the expected estimates derived by local misspecification. The green lines give the least false estimates. The horizontal line is at the true value. The left plot is for θ_0 and the right for θ_1.
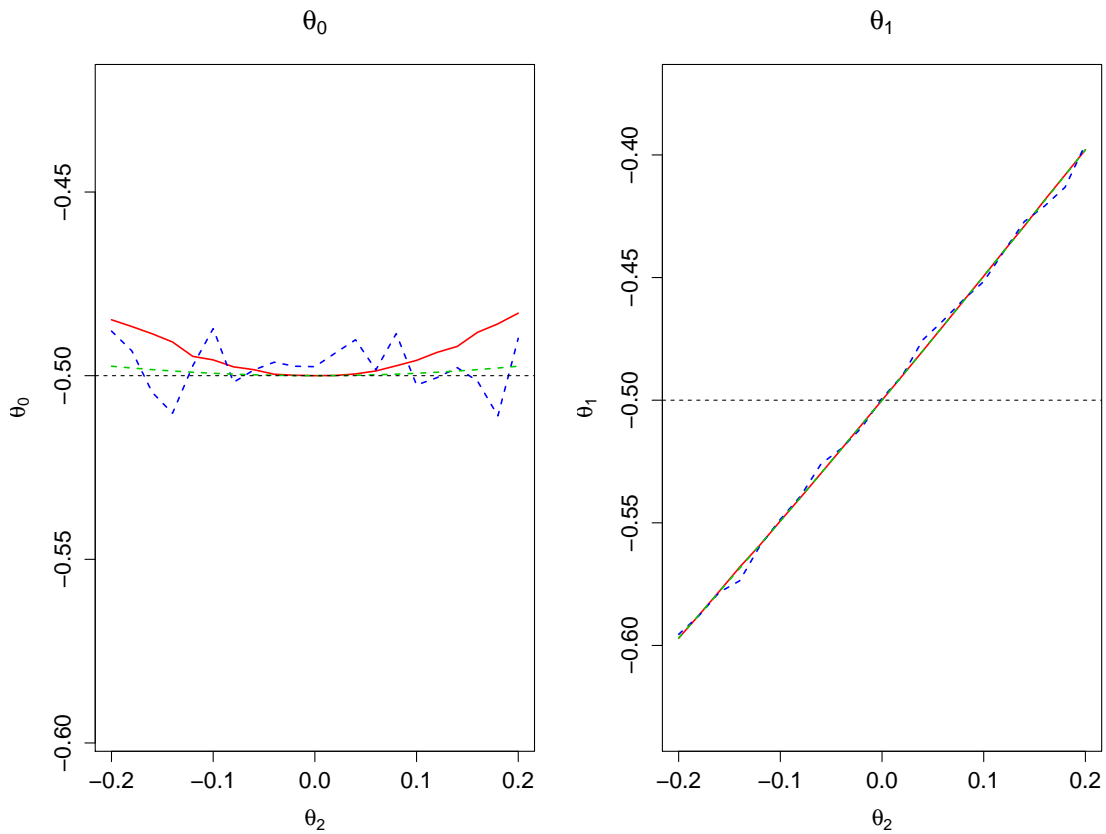
Figure 8.2: Parameters $\theta = (1, -1)$. Lines as Figure 8.1.
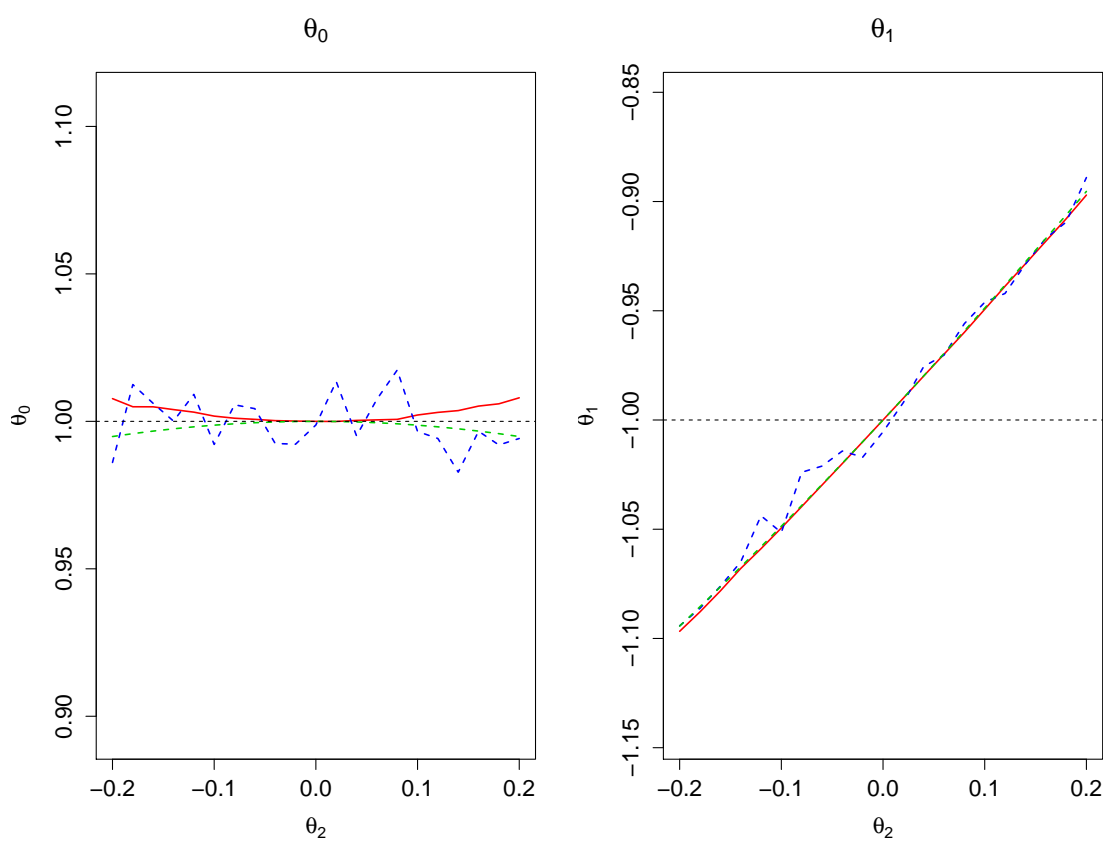
8.3.1 CI coverage for the estimated θ_0 and θ_1

Here we calculate the 95% confidence interval (CI) coverage percentages for the parameters of interest θ_0 and θ_1 by using the three different estimation methods:

- The General Linear Model estimation method (*glm*). The logistic regression parameters are estimated using *glm* method and they are denoted by θ_0^{Raw} and θ_1^{Raw}, which mean raw results or unadjusted.

- The local misspecification method using equation (8.6), the parameters here are denoted by θ_0^{**} and θ_1^{**}.

- The least false method in Chapter 4, the parameters associated with this method are denoted by θ_0^* and θ_1^*.

For the local misspecification and least false methods, we obtain the CI by shifting the location by the estimated bias. We will perform three different comparisons. In Table 8.1, we assume θ_2=0, so MAR is correct here. In Table 8.2 and 8.3 we assume MNAR with assumed θ_2 (denoted by θ_2^A) =+0.1 and -0.1 respectively. All are based on 1000 simulation repetitions each of sample size 1000. We calculate the CI coverage under a range of true θ_2 (denoted by θ_2^T) from -0.2 to +0.2.

In Table 8.1, we note that all the methods give the same coverage for each of θ_0 and also for θ_1 in each row (at the same true θ_2). For example, the estimated CI coverage for θ_0 in the third, fifth and seventh columns are identical, also the fourth, sixth and eighth columns are identical. In addition, the confidence interval coverages are good for θ_0 at all true θ_2, because of the stability of the estimates as seen in Figures 8.1 and 8.2. For θ_1, when the MAR model is correct (θ_2^T=0) all methods perform well, but when it is not there is serious undercoverage. The reason is that under MAR there is no bias and so all the methods are supposed to give similar estimates.

Table 8.1: The CI coverage at assumed $\theta_2=0$ in %.

θ_2^T	θ_2^A	θ_0^{Raw}	θ_1^{Raw}	θ_0^{**}	θ_1^{**}	θ_0^{*}	θ_1^{*}
-0.20	0.00	95.10	39.80	95.10	39.80	95.10	39.80
-0.18	0.00	96.20	50.10	96.20	50.10	96.20	50.10
-0.16	0.00	95.90	59.20	95.90	59.20	95.90	59.20
-0.14	0.00	95.30	64.60	95.30	64.60	95.30	64.60
-0.12	0.00	96.00	76.00	96.00	76.00	96.00	76.00
-0.10	0.00	95.50	78.40	95.50	78.40	95.50	78.40
-0.08	0.00	95.80	84.30	95.80	84.30	95.80	84.30
-0.06	0.00	94.80	89.40	94.80	89.40	94.80	89.40
-0.04	0.00	95.20	93.30	95.20	93.30	95.20	93.30
-0.02	0.00	95.80	95.30	95.80	95.30	95.80	95.30
0.00	0.00	95.90	95.60	95.90	95.60	95.90	95.60
0.02	0.00	95.70	95.30	95.70	95.30	95.70	95.30
0.04	0.00	96.10	92.10	96.10	92.10	96.10	92.10
0.06	0.00	95.80	87.60	95.80	87.60	95.80	87.60
0.08	0.00	95.90	80.80	95.90	80.80	95.90	80.80
0.10	0.00	96.00	74.80	96.00	74.80	96.00	74.80
0.12	0.00	95.10	61.90	95.10	61.90	95.10	61.90
0.14	0.00	93.80	53.50	93.80	53.50	93.80	53.50
0.16	0.00	95.00	37.70	95.00	37.70	95.00	37.70
0.18	0.00	96.50	31.10	96.50	31.10	96.50	31.10
0.20	0.00	94.50	21.80	94.50	21.80	94.50	21.80

Table 8.2: CI coverage at assumed $\theta_2 = 0.1$ in %.

θ_2^T	θ_2^A	θ_0^{Raw}	θ_1^{Raw}	θ_0^{**}	θ_1^{**}	θ_0^*	θ_1^*
-0.20	0.10	95.40	36.80	95.20	7.00	95.20	6.90
-0.18	0.10	94.80	52.40	94.40	9.40	94.20	9.10
-0.16	0.10	95.70	56.00	95.50	14.10	95.40	13.80
-0.14	0.10	95.50	65.80	95.60	20.70	95.50	20.30
-0.12	0.10	95.30	73.20	95.30	27.60	95.20	27.00
-0.10	0.10	95.90	80.80	95.90	34.60	95.80	34.20
-0.08	0.10	95.60	88.00	95.80	44.20	95.70	43.50
-0.06	0.10	94.60	88.90	94.30	53.70	94.20	52.60
-0.04	0.10	94.70	94.30	95.10	61.80	95.10	61.60
-0.02	0.10	95.90	94.40	95.70	65.70	95.80	65.40
0.00	0.10	95.70	96.30	95.80	74.60	95.60	74.40
0.02	0.10	95.20	93.20	94.90	84.70	94.90	84.20
0.04	0.10	95.70	92.60	95.50	89.20	95.60	88.60
0.06	0.10	96.40	88.50	96.70	93.50	96.60	93.30
0.08	0.10	94.50	81.30	94.50	94.70	94.50	94.60
0.10	0.10	95.50	73.00	95.60	95.80	95.60	95.60
0.12	0.10	96.00	61.40	96.20	93.40	96.10	93.50
0.14	0.10	95.70	52.70	95.30	90.80	95.30	90.80
0.16	0.10	95.10	41.60	95.30	88.50	95.30	88.60
0.18	0.10	95.10	29.70	94.80	80.00	94.80	80.30
0.20	0.10	94.90	22.20	95.00	69.20	95.10	69.60

Table 8.2 provides an example for the CI coverages when the MNAR assumption is valid, and we assume $\theta_2 = +0.1$. The coverages of θ_0^{Raw}, θ_0^{**} and θ_0^* are good at any true θ_2 in the range (-0.2, +0.2). Turning to the coverage of different estimates of θ_1, the *glm* estimate, θ_1^{Raw}, gives good coverage (about 95%) only at true $\theta_2 = 0$. However, for θ_1^{**} and θ_1^*, both corrections work very well when the assumed θ_2 matches the true value (0.1) as expected (no bias). However, when the assumed θ_2 is wrong, the coverage of θ_1^{**} and θ_1^* decreases dramatically and reach only around 7% at true $\theta_2 = -0.2$. The poor coverage rates of the nominal confidence intervals are due to the bias and not to poor variance estimation, argued Robins et al. [1995].

When the assumed θ_2 is -0.1, Table 8.3 leads to similar conclusions as in Table 8.2. As assumed θ_2 has opposite sign of the assumed θ_2 in the previous table, the most undercovarges now are associated with the positive values of true θ_2. Again, θ_0 estimators give good coverage for all true θ_2 in (-0.2,+0.2). Turning to the θ_1 estimators, the CI coverage for θ_1^{Raw} is good (96.20%) when data is really MAR (at $\theta_2^T = 0$). On the other hand, θ_1^{Raw} CI coverage decreases as true θ_2 goes further from zero in both directions, as seen before in Table 8.2. Regarding the CI coverages for the estimates θ_1^{**} and θ_1^{*}, they both have good coverage only when $\theta_2^T = \theta_2^A$ (at true $\theta_2 = -0.1$), but otherwise they have serious undercoverage.

Table 8.3: CI coverage at assumed θ_2=-0.1 in %.

θ_2^T	θ_2^A	θ_0^{Raw}	θ_1^{Raw}	θ_0^{**}	θ_1^{**}	θ_0^{*}	θ_1^{*}
-0.20	-0.10	95.80	41.90	95.90	82.00	96.00	81.00
-0.18	-0.10	95.50	47.70	95.60	88.00	95.50	87.80
-0.16	-0.10	96.00	55.70	96.20	91.00	96.00	91.00
-0.14	-0.10	97.10	66.00	97.10	94.20	97.00	93.90
-0.12	-0.10	94.90	72.20	94.70	95.10	94.70	94.90
-0.10	-0.10	95.00	79.90	95.00	95.40	95.00	95.40
-0.08	-0.10	94.80	86.90	94.80	93.60	94.80	93.40
-0.06	-0.10	95.50	90.60	95.10	91.90	95.10	91.90
-0.04	-0.10	96.00	93.90	95.70	90.60	95.80	90.90
-0.02	-0.10	96.00	94.50	96.10	83.00	96.10	83.80
0.00	-0.10	94.40	96.20	94.20	77.10	94.20	77.50
0.02	-0.10	95.50	93.50	95.50	65.70	95.40	66.20
0.04	-0.10	95.60	92.00	95.50	58.70	95.50	59.10
0.06	-0.10	94.90	87.20	95.00	47.20	94.70	47.40
0.08	-0.10	95.50	81.50	95.30	34.50	95.30	35.10
0.10	-0.10	95.00	75.20	95.40	29.20	95.20	29.40
0.12	-0.10	96.40	67.30	96.30	21.00	96.20	21.70
0.14	-0.10	95.00	51.10	94.70	12.10	94.80	12.40
0.16	-0.10	95.80	41.40	95.70	7.10	95.50	7.30
0.18	-0.10	95.90	32.90	95.90	4.80	96.00	4.80
0.20	-0.10	95.60	22.70	95.10	2.60	94.90	2.60

8.3.2   Sensitivity analysis using simulated data

We simulate two data sets of sample size 1000, one with true $\theta_2=0$ assuming MAR and the other with true $\theta_2=0.5$ (MNAR). In both data sets we use the dropout parameters $\theta_0 = -0.5$ and $\theta_1 = -0.5$. We use a grid of assumed θ_2 from -1.0 to 1.0 to estimate the model coefficients (θ_0, θ_1) using the local misspecification and least false methods. Next the confidence intervals for the estimates of $\theta = (\theta_0, \theta_1)$ are adjusted for the various assumed θ_2. Figures 8.3 and 8.4 illustrate. The blue lines are when we use local misspecification for adjustment, the red lines are for the least false values and the horizontal lines are at the true values (at $\theta_0 = -0.5$ and $\theta_1 = -0.5$).

Figure 8.3: The 95% CI of θ_0 and θ_1 under MAR. We use $\theta_0 = -0.5$, $\theta_1 = -0.5$ and true $\theta_2 = 0$. The blue lines are local misspecification, the red for least false and the horizontal lines are at the MAR maximum likelihood estimates.
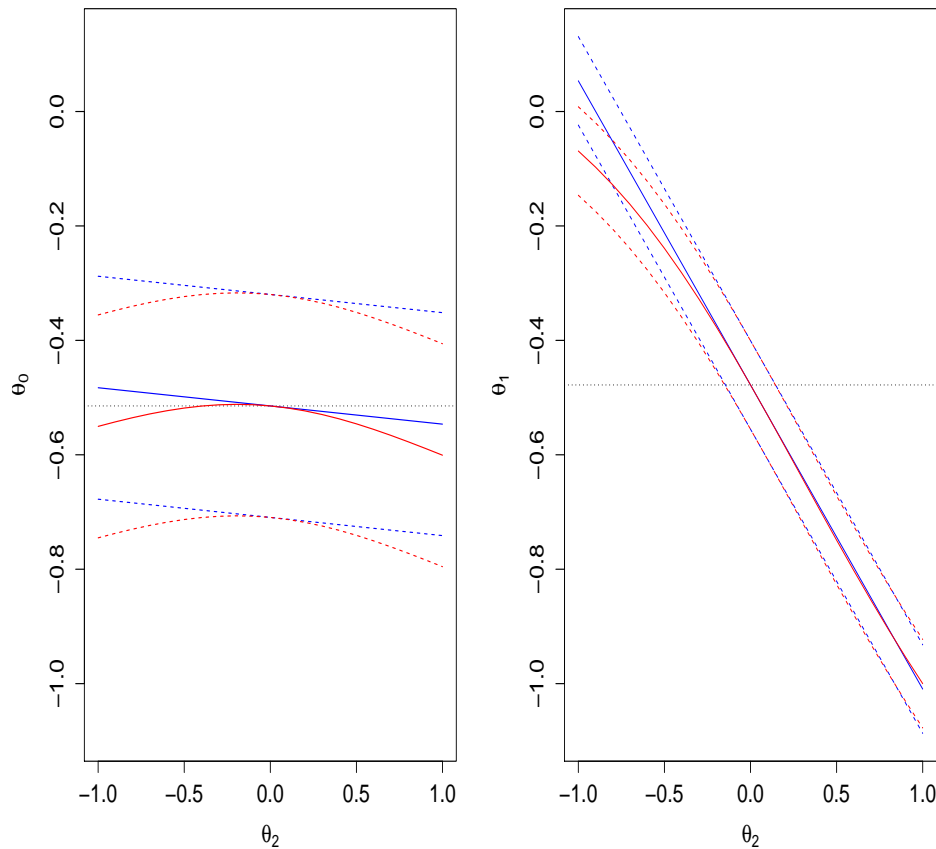
The left plot of Figure 8.3 is for θ_0, the right for θ_1. Similarly to the results found in Table 8.1, the left plot shows that θ_0 is not substantially affected by the different assumed values of θ_2. In contrast, the right plot shows that θ_1 is highly affected by changes in assumed θ_2. Local misspecification and least false values estimates are very similar in how they respond to the change of assumed θ_2 in both plots, the lines are close. However, there is

some curvature for low θ_2 under least false but not local misspecification as shown in the right plot.

Figure 8.4: The 95% CI of θ_0 and θ_1 under MNAR. We use $\theta_0 = -0.5$, $\theta_1 = -0.5$ and true $\theta_2 = 0.5$. Lines as in Figure 8.3.
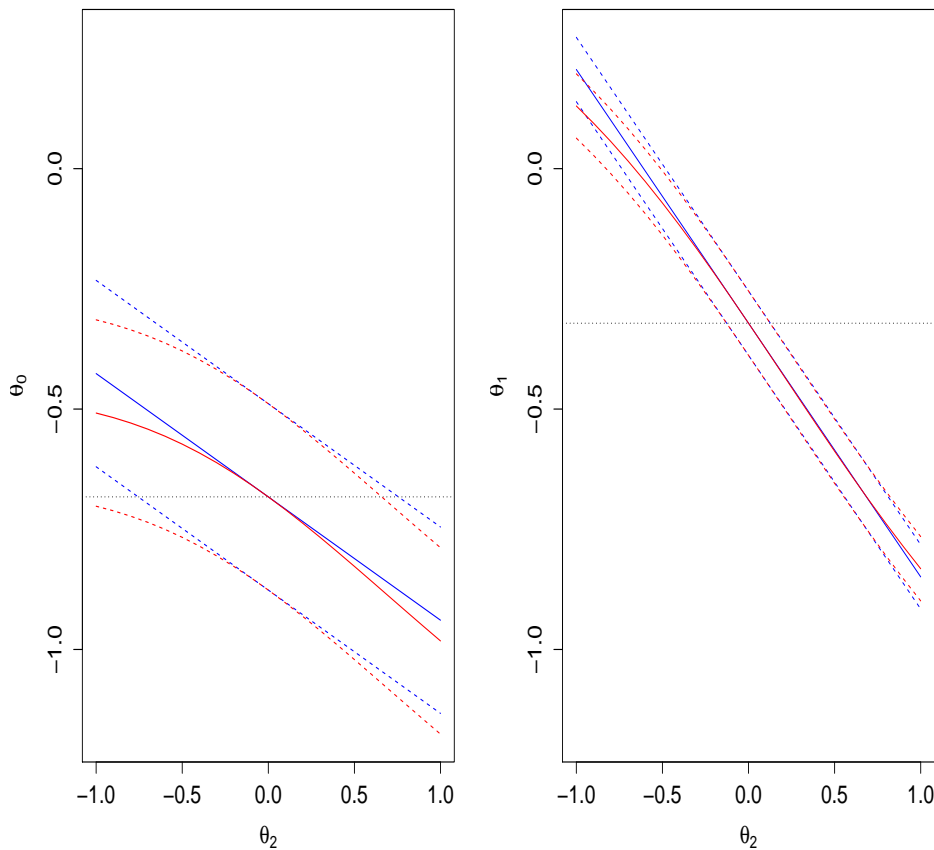
Figure 8.4 is obtained when the data really are MNAR. To compare with Figure 8.3, the left plot in Figure 8.4 shows that the θ_0 estimates are more sensitive to the variation of assumed θ_2, as the slope of the lines is clearly greater. Again, the least false estimates for θ_0 and θ_1 show curvature for low θ_2. Note that at assumed $\theta_2 = 0$ the horizontal line and the estimates lines cross. For example, in the right plot, at assumed $\theta_2 = 0$, the lines cross at around $\theta_1 = -0.3$, which is the maximum likelihood estimate of the dropout parameter θ_1 under MAR. Clearly this estimate is biased due to the misspecification of fitting MAR to MNAR data. At $\theta_2 = 0$ the CIs in each plot do not include the true value of θ_0 and θ_1 respectively.

8.4   Application

8.4.1   Sensitivity analysis: Two timepoints example

In this section we use the two timepoints real data example, which is the same data as used in Chapter 7. There are two treatments and two measurement times. The covariates are only treatment type and time. There are 422 participants, assigned to either Treatment A or B. At time 1, all subjects provided a response, but 24.4% dropped out at time 2.

A sensitivity analysis approach over a grid of assumed θ_2 from -1.5 to +1.5 using the local misspecification and least false method is shown in Figure 8.5. The left plot shows confidence intervals for θ_0 as the assumed value of θ_2 varies. The right plot shows the confidence intervals for θ_1 for different assumed θ_2. The blue lines are local misspecification (estimates and the 95% CI) and the red are least false. The horizontal line is the estimate under MAR.

Figure 8.5: The 95% CI for θ_0 and θ_1: Two timepoints example. The left plot refers to θ_0 and the right is for θ_1. The blue lines are local misspecification estimates and the red lines are the least false. The horizontal line is the estimate under MAR.

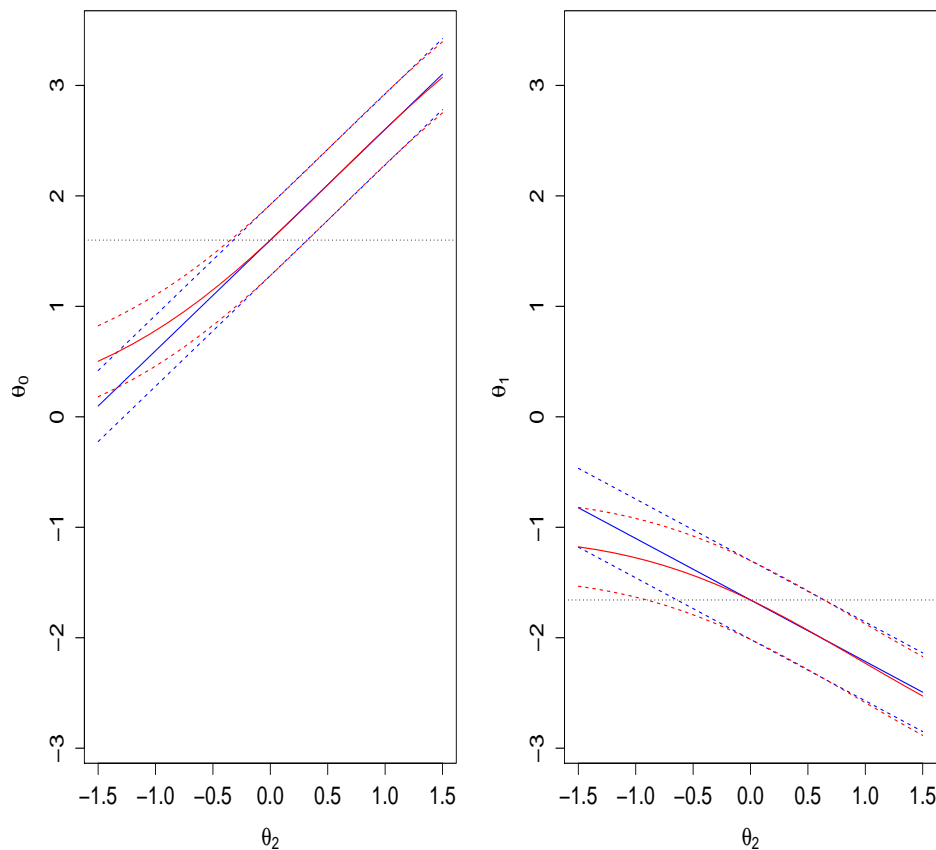Figure 8.5 shows how we should adjust the confidence intervals for θ_0 and θ_1, obtained

under MAR, under the assumption that θ_2 is not zero. The MAR estimates are 1.60 and -1.66 for θ_0 and θ_1 respectively, and so θ_2 in +/- 1.5 is believable. The local misspecification bias adjustment seems linear in this example. Least false has some curvature at the lowest θ_2 but is otherwise very close to local misspecification result. The plot is qualitatively similar to Figure 8.4, which used simulated data. Note that the CI width for θ_0 and θ_1 is approximately equal. Also, both of the estimates go further from zero as θ_2 increases.

Table 8.4: Summary of MAR logistic regression coefficients θ_0 and θ_1: Two timepoints example.

Coefficient	Estimate	s.e	z value
θ_0	1.60	0.16	9.95
θ_1	-1.66	0.18	-9.30

Results obtained from Table 8.4 match the results of Figure 8.5. The standard error (s.e) for θ_1 equals 0.18 which is almost equal to the standard error for θ_0=0.16. The z values for both θ_0 and θ_1 indicate that the values of θ_0 and θ_1 are significantly different from zero.

8.4.2 Sensitivity analysis: Three timepoints example

In this section we use the same three timepoints data as used in Chapter 7. This data (schizrenia) is from a clinical trial with three treatment groups and three measurement times. As mentioned before in Chapter 7, the trial compared three treatments: A standard therapy (Treatment 1), a placebo (Treatment 2), and an experimental therapy (Treatment 3). The response is PANSS (positive and negative symptom score, a measure of mental health, with high values being bad). Values ranged from 30 to 210. Typically we expect a schizophrenia patient in this clinical trial to have score around 90. Of the 518 patients, 249 dropped out and the missingness percentages are around 24% and 48% at the second and third times respectively. We assume that the dropout depends on the response Y (PANSS) but not treatment groups. The missingness according to the groups shows that treatment group 2 has the highest dropout percentage while treatment groups 1 and 3 have lower dropout percentages.

The MAR dropout model is (1.6) as defined in the Introduction.

Figure 8.6: The 95% CI for θ_0 and θ_1 between time 1 and 2: Three timepoints example. Lines as Figure 8.5.

Dropout between time 1 and time 2
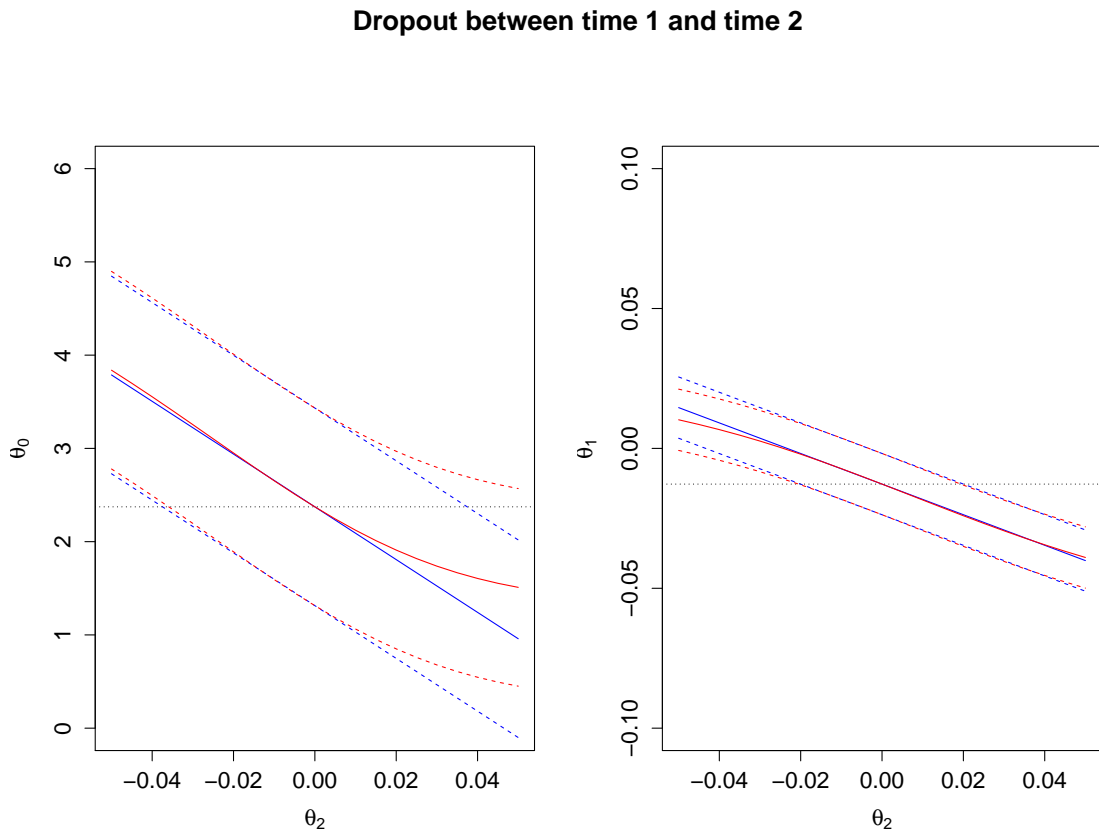
Figure 8.6 shows the 95% CI for the estimates of θ_0 and θ_1 between time 1 and 2 and Figure 8.7 is for CI of estimates between time 2 and 3. We take a grid of assumed θ_2 from -0.05 to +0.05, since the estimate under MAR is θ_1=-0.01 for dropout between time 1 and 2, and -0.03 for dropout between time 2 and 3. The horizontal line is the estimate under MAR. The red lines are the least false values (estimates and the 95% CI) and the blue lines are the local misspecification estimates.

Figure 8.7: The 95% CI for θ_0 and θ_1 between time 2 and 3: Three timepoints example. Lines as Figure 8.5.

Dropout between time 2 and time 3

Table 8.5: Summary of MAR logistic regression coefficients: Three timepoints example, data between time 1 and 2

Coefficient	Estimate	s.e	z value
θ_0	2.37	0.53	4.48
θ_1	-0.01	0.01	-2.33

In Table 8.5, the results obtained match Figure 8.6. The standard error (s.e) for θ_0 =0.53 which is greater than the standard error for θ_1=0.01, which is clearly seen in the figure as θ_0 has wider CI than θ_1. The z values for both θ_0 and θ_1 indicate that the values of θ_0 and θ_1 are significantly different from zero.
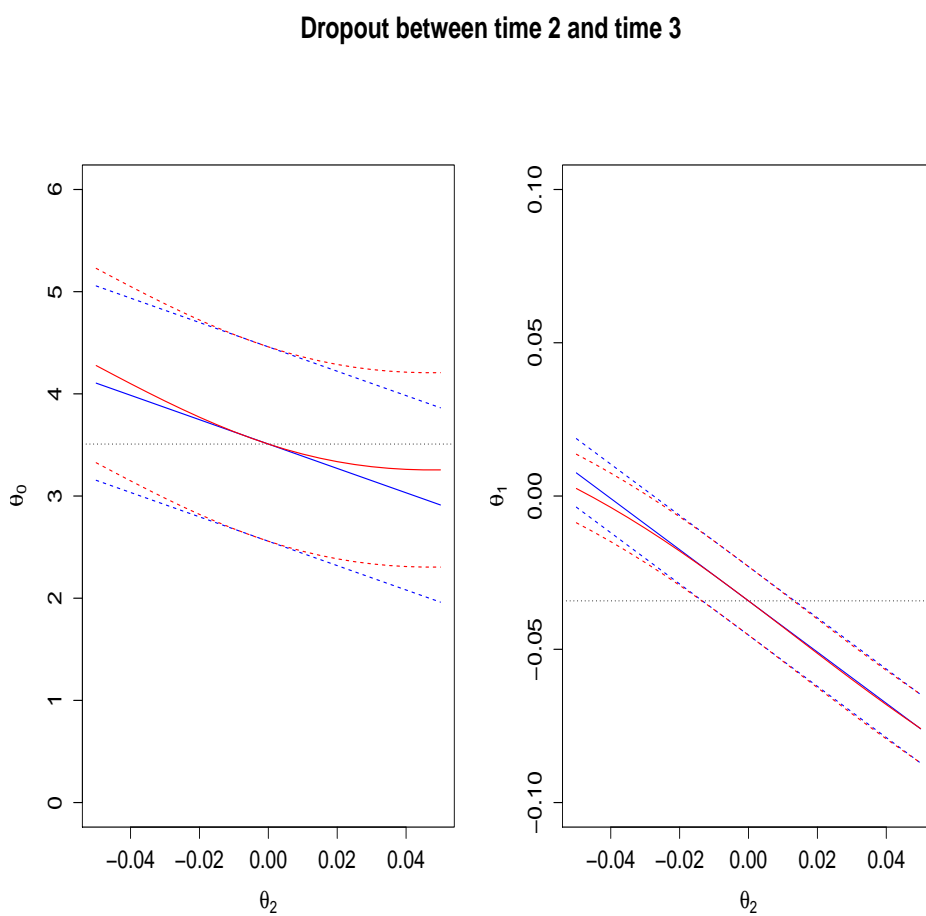
Table 8.6: Summary of MAR logistic regression coefficients: Three timepoints example, data between time 2 and 3

Coefficient	Estimate	s.e	z value
θ_0	3.51	0.48	7.38
θ_1	-0.03	0.01	-6.12

From Table 8.5 the value of $\hat{\theta}_0$ between time 1 and time 2 under MAR is 2.4, while $\hat{\theta}_1$=-0.01. The dropout probability estimates under MAR between time 2 and 3, as shown in Table 8.6, are $\hat{\theta}_0$=3.5, $\hat{\theta}_1$=-0.03. The results obtained from Table 8.6 match Figure 8.7. The standard error (s.e) for θ_1 is 0.01 which is less than the standard error for θ_0=0.48, that is clear in the figure as θ_0 has wider CI than θ_1. The z values for both θ_0 and θ_1 indicate that the values of θ_0 and θ_1 are significantly different from zero.

8.5 Local Misspecification by IPW Method

The aim in this section is to estimate the parameters of the estimating equation (GEE) i.e. β's by using the IPW method which depends on the results obtained in the first part of this chapter. We use Copas and Eguchi method again to get the local misspecification estimates. Similar to the approach used in Chapter 7, again in this chapter we compare the estimates based on local misspecification with the simulation estimates. We will not use the least false values derived through the IPW method which we calculated in Chapter 4 because there is no closed form for β_3^* and β_4^* under MNAR but instead they are obtained by numerical calculations, as mentioned in Chapter 4. We know from Chapter 2 that the IPW works under missing at random (MAR) so we will consider the working model is MAR while the truth is MNAR.

8.5.1 Simulation study

We use our familiar notation. For comparison, we first compute the β_3 estimates and the β_4 estimates using the local misspecification method and IPW estimation method in simulations. We conduct four simulation examples that differed in the choice of the parameters. Each simulation experiment was based on 20 repetitions each of size 1000. We take the average for the estimates over these runs. Figures 8.8, 8.9, 8.10 and 8.11 show mean simulation estimates and mean local misspecification approximations for four examples, each of which is calculated over a grid of θ_2 in the range (-0.2,0.2). The blue lines (dotted lines) are the mean estimated values of β_3 and β_4 under a MAR assumption when the data are really MNAR using simulations. The red lines (solid lines) show mean values derived from the local misspecification method.

The main objective of these figures is to reassure us that the local misspecification method gives estimates matching those that are obtained using the IPW method. The idea is we know that the IPW is consistent under MAR and so this estimating equation has to work. We only use the observed subjects at time 2. The results obtained here are similar to what we found in the equivalent figures in Chapter 7. There are some differences in the estimates but they both have the same magnitude. For example, in comparing Figure 8.8 here with Figure 7.1 in Chapter 7, where we used the same parameter set used here, it is clear that the estimates of β_3 increase as the assumed value of θ_2 increases. Also, in Figure 7.1 in Chapter 7 and Figure 8.8 here in this chapter, the β_4 estimate decreases as θ_2 increases. And again, the two lines cross at $\theta_2=0$, which means MAR data. Generally the local misspecification method matches the estimated IPW values, though there are some differences for the most extreme θ_2.

Figure 8.8: Parameters $\beta = (-2, -2, -1, -1)$, $\theta = (-0.5, -0.5)$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1$, $\rho = 0.5$. This gives dropout rate $\approx 40\%$. The blue lines are the average estimates over 20 simulations using IPW estimation. The red lines are the local misspecification estimates. The horizontal lines are at the true values.
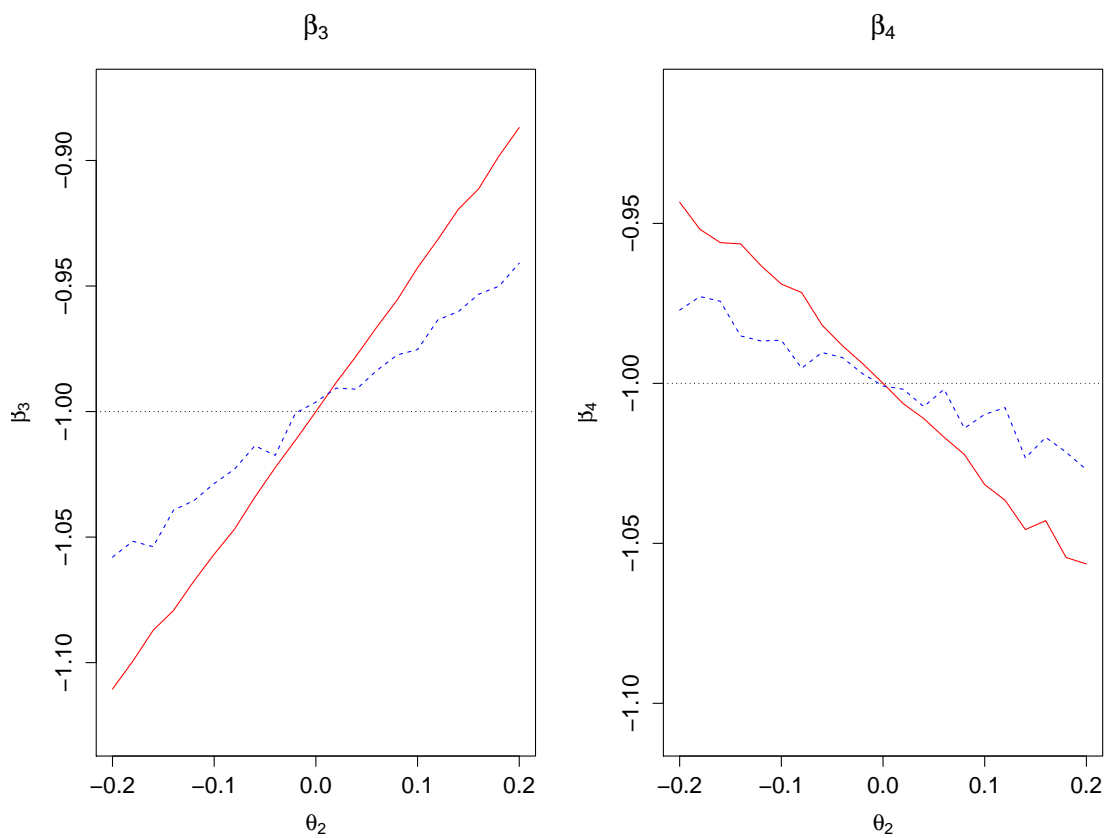
Figure 8.9: Parameters $\beta = (-1, -1, 1, 2)$, $\theta = (-0.5, -0.5)$, $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, $\rho = 0.5$. This gives dropout rate $\approx 50\%$. Lines as in Figure 8.8

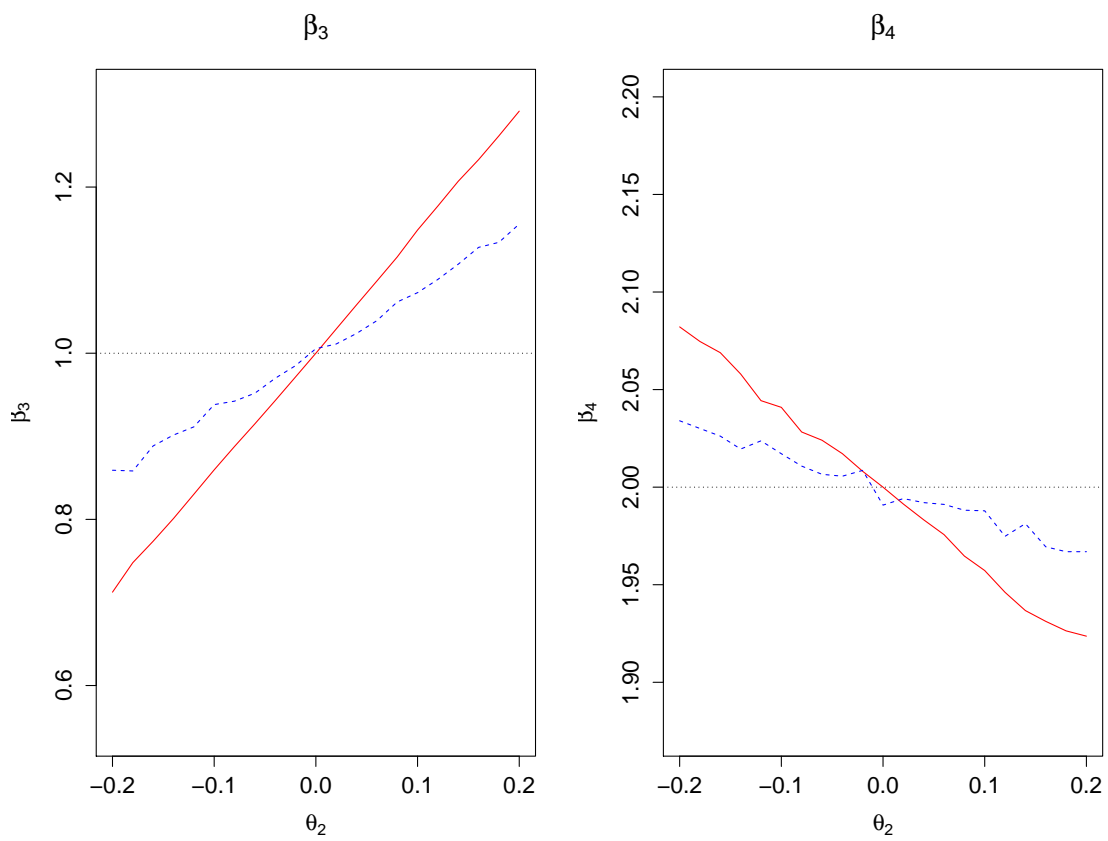Figure 8.10: The parameter set used here is $\beta = (-2, -2, -1, -1)$, $\theta = (-1, 0.5)$, $\sigma_1^2 = 2$, $\sigma_2^2 = 1$, $\rho = 0$. This gives dropout rate $\approx 83\%$. Lines as in Figure 8.8
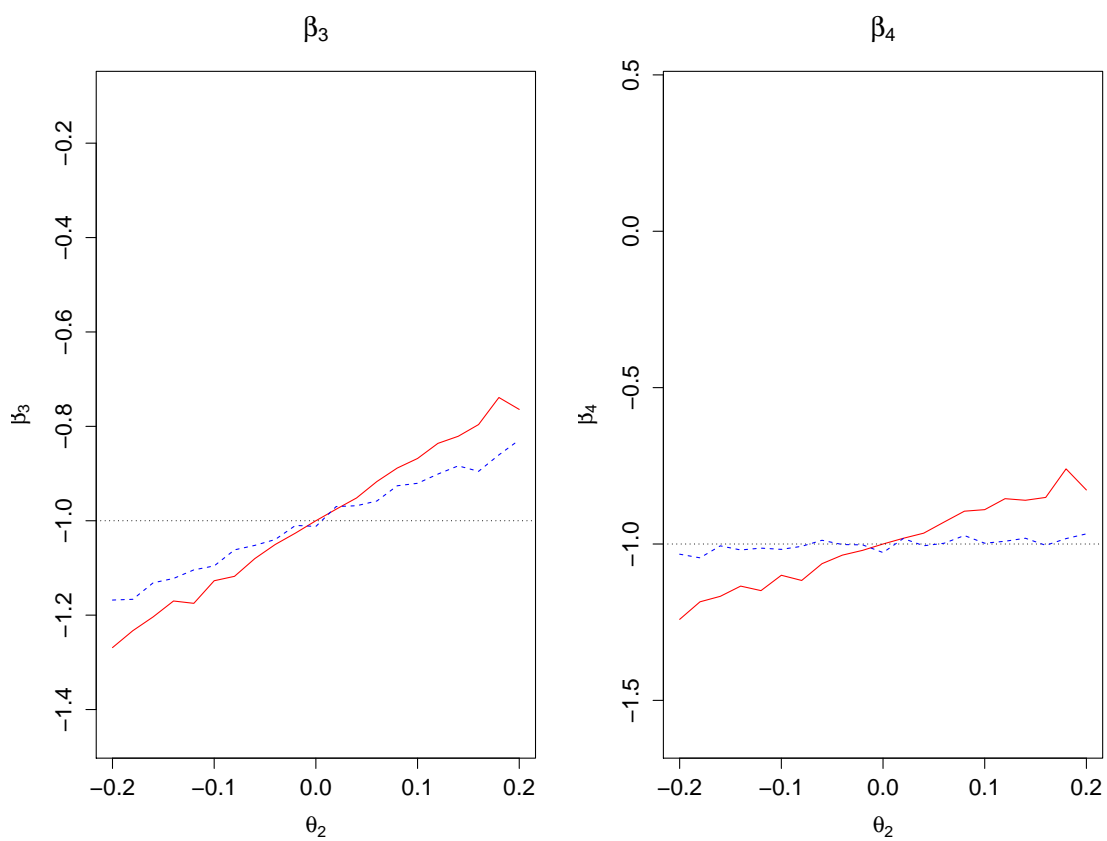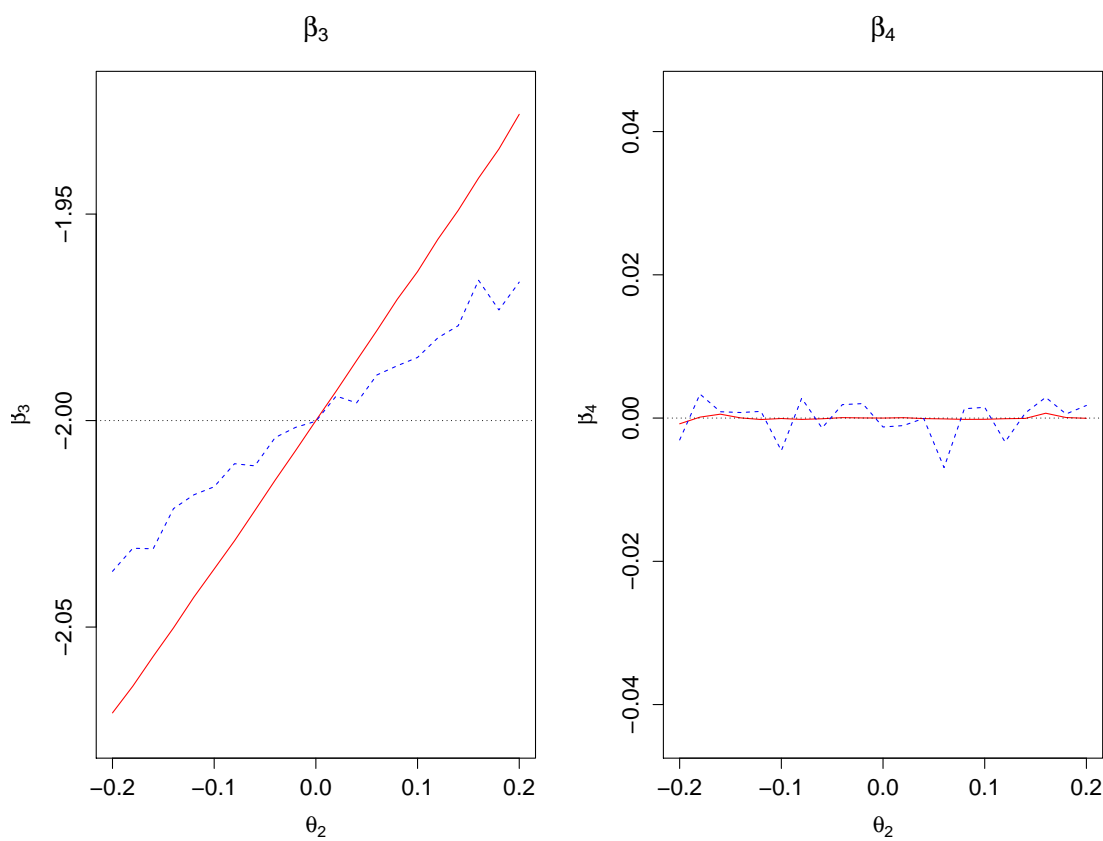
Figure 8.11: We use $\beta = (-2, 0, -2, 0)$, $\theta = (-0.5, -0.5)$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1$, $\rho = 0.75$, and this set leads to dropout rate $\approx 48\%$. Lines as in Figure 8.8

8.5.2   The CI coverage for the estimated β_3 and β_4

Here we calculate the 95% CI coverage for the local misspecification estimates obtained using the IPW method, which are denoted by β_3^{**} and β_4^{**}. We present three different examples, we use assumed $\theta_2=0$ in Table 8.7, assumed $\theta_2=-0.05$ in Table 8.8 and assumed $\theta_2=+0.05$ in Table 8.9. Each simulation experiment was based on 1000 repetitions at sample size 1000. Table 8.7 reports the coverage probability of nominal 95% confidence

Table 8.7: The CI coverage at assumed $\theta_2=0$ in %.

θ_2^T	θ_2^A	β_3^{**}	β_4^{**}
-0.10	0.00	90.50	90.70
-0.09	0.00	90.10	94.10
-0.08	0.00	91.90	91.20
-0.07	0.00	93.40	94.10
-0.06	0.00	93.60	94.10
-0.05	0.00	94.10	92.50
-0.04	0.00	94.80	94.70
-0.03	0.00	94.30	93.00
-0.02	0.00	94.30	93.90
-0.01	0.00	94.70	94.10
0.00	0.00	95.90	94.40
0.01	0.00	96.70	94.70
0.02	0.00	95.60	94.60
0.03	0.00	94.90	94.50
0.04	0.00	95.10	94.20
0.05	0.00	94.00	94.10
0.06	0.00	92.00	94.60
0.07	0.00	94.20	92.80
0.08	0.00	93.80	94.40
0.09	0.00	93.90	94.70
0.10	0.00	93.30	94.40

intervals, for the estimates of β_3^{**} and β_4^{**} when assuming MAR. Since the estimators of β_3 and β_4 at true $\theta_2=0$ were unbiased, thus the coverage probabilities are very close to the nominal level of 95%. Otherwise, there is slight undercoverage for the CIs of β_3 and β_4 when the true θ_2 differs from zero.

Table 8.8: The CI coverage at assumed θ_2=-0.05 in %.

θ_2^T	θ_2^A	β_3^{**}	β_4^{**}
-0.10	-0.05	95.00	92.90
-0.09	-0.05	96.00	94.40
-0.08	-0.05	95.20	93.60
-0.07	-0.05	93.50	92.80
-0.06	-0.05	94.00	93.20
-0.05	-0.05	94.80	93.30
-0.04	-0.05	94.20	93.40
-0.03	-0.05	93.10	95.10
-0.02	-0.05	90.60	92.80
-0.01	-0.05	92.20	93.00
0.00	-0.05	91.50	93.60
0.01	-0.05	91.30	92.40
0.02	-0.05	87.90	92.20
0.03	-0.05	91.10	93.50
0.04	-0.05	87.20	94.20
0.05	-0.05	86.70	92.50
0.06	-0.05	84.90	93.60
0.07	-0.05	81.80	91.80
0.08	-0.05	83.50	93.20
0.09	-0.05	79.50	91.40
0.10	-0.05	80.80	92.00

Table 8.8 shows that unless true θ_2=-0.05 the estimators β_3^{**} and β_4^{**} are biased, and this bias leads to undercoverage. For example, at true θ_2=0.10, the coverage of β_3^{**} decreased roughly by 14%. Also, note that β_3^{**} has a decreasing coverage as true θ_2 increases, while in Table 8.9, β_3 has the opposite pattern, i.e. the coverage increases as true θ_2 increases.

In conclusion, if true θ_2 and assumed θ_2 are the same (so we assumed right) then we expect 95% coverage. If assumed θ_2 is a long way from true θ_2 then coverage is expected to be poor. However, β_3 seems to be more affected than β_4.

Table 8.9: The CI coverage at assumed θ_2=0.05 in %.

θ_2^T	θ_2^A	β_3^{**}	β_4^{**}
-0.10	0.05	77.40	89.60
-0.09	0.05	78.70	91.40
-0.08	0.05	82.60	90.60
-0.07	0.05	81.30	90.70
-0.06	0.05	83.30	91.00
-0.05	0.05	86.70	92.40
-0.04	0.05	87.10	92.70
-0.03	0.05	88.10	93.80
-0.02	0.05	88.80	91.80
-0.01	0.05	90.50	93.30
0.00	0.05	91.10	93.40
0.01	0.05	93.30	93.90
0.02	0.05	92.90	92.90
0.03	0.05	93.10	95.50
0.04	0.05	94.00	94.30
0.05	0.05	94.70	94.00
0.06	0.05	95.30	94.30
0.07	0.05	95.80	94.20
0.08	0.05	95.00	94.30
0.09	0.05	95.80	93.10
0.10	0.05	96.50	94.40

8.5.3   Sensitivity analysis using simulated data

In this section we perform a sensitivity analysis to examine how our inferences concerning the regression parameters β change as θ_2 varies over a range of (-0.1,+0.1). In Figure 8.12, MAR is correct and the unadjusted confidence intervals (red lines) include the true parameter values (-1 and -1), as in this case so do the adjusted ones (blue lines). It is clear that the CIs for β_4 are wider and this is consistent with our finding in the earlier tables where the CI coverages for β_4 were greater than the β_3 CI coverage. In Figure 8.13 we have fitted MAR to data that are really MNAR. The best estimates, as expected, are at the true θ_2. Of course we would not know that in practice, as mentioned in Chapter 7. This figure is equivalent to Figure 7.8 in Chapter 7.

Figure 8.12: The 95% CI of β_3 and β_4 under MAR. We use β_3=-1, β_4=-1, $\theta_0 = -0.5$, $\theta_1 = -0.5$ and $\theta_2 = 0$. The red lines are the unadjusted confidence intervals, and the adjusted ones are the blue lines.
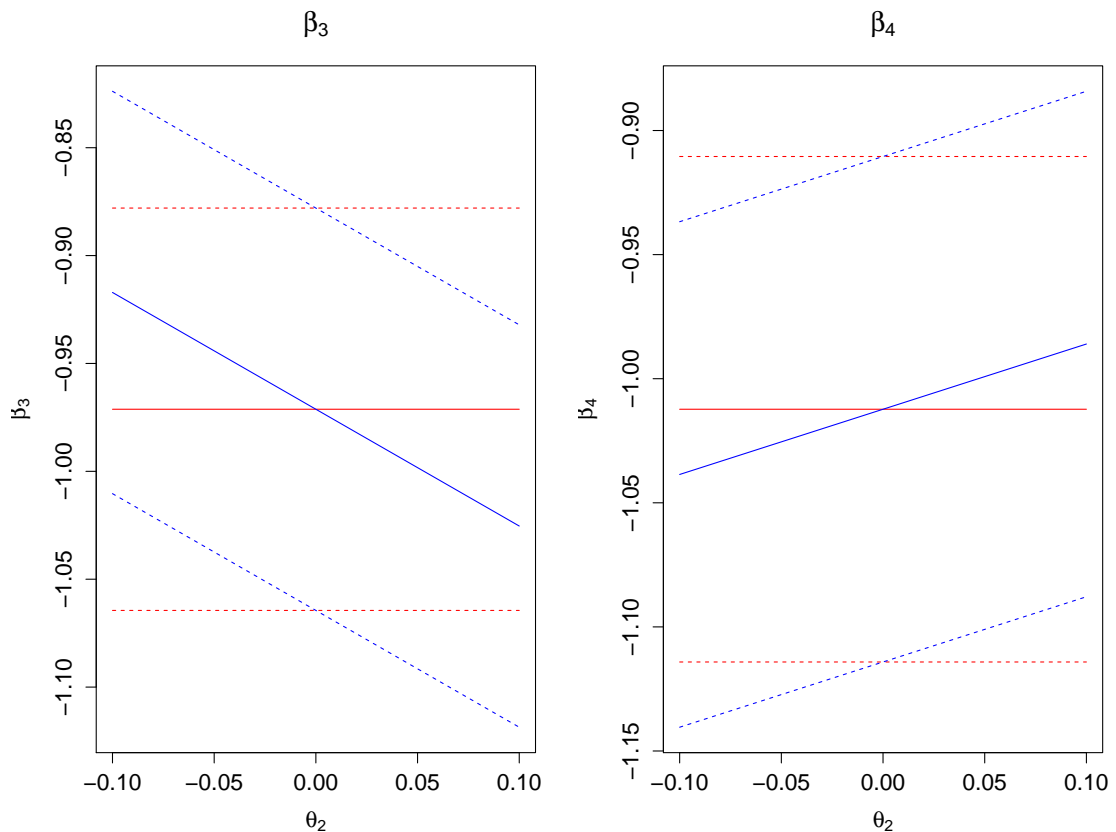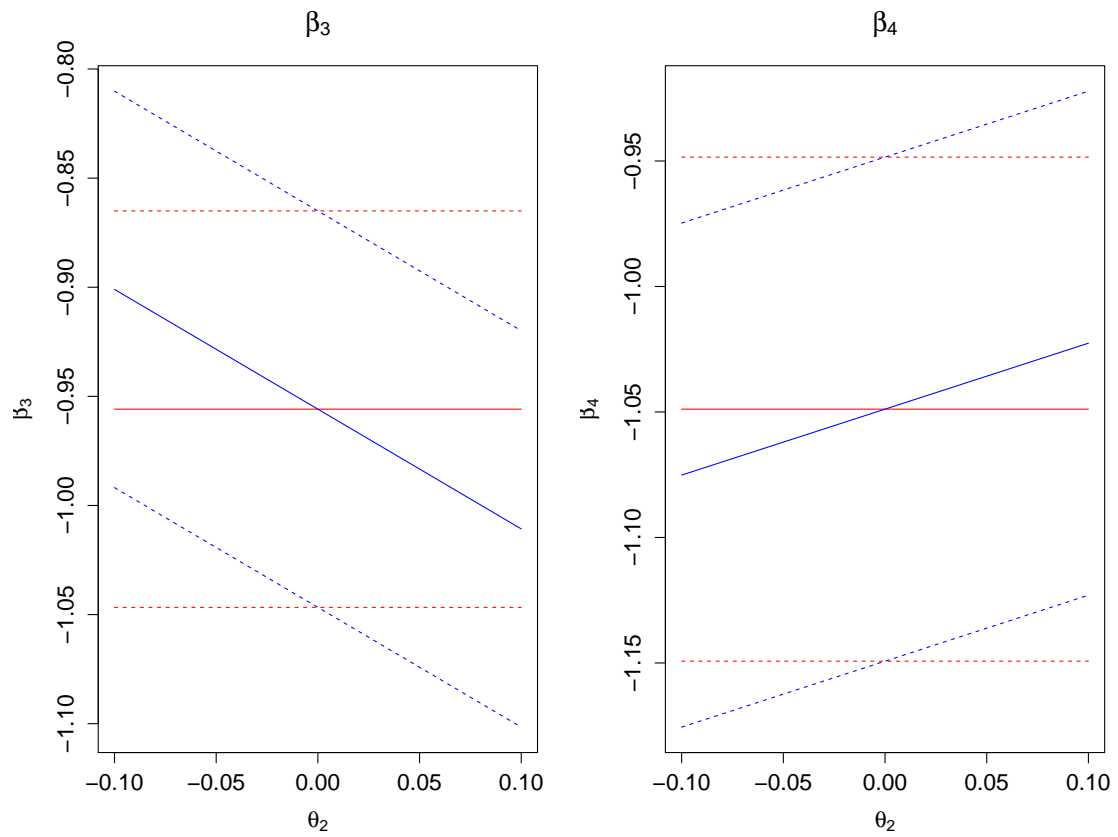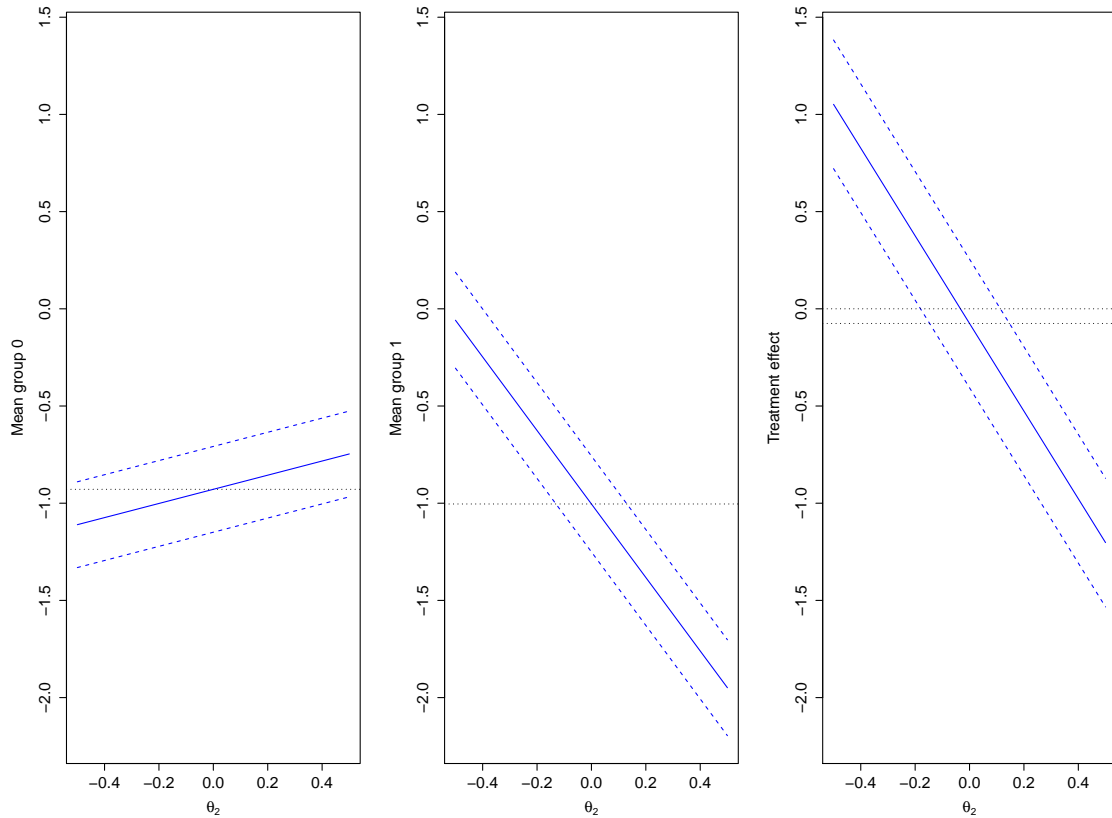
Figure 8.13: The 95% CI of β_3 and β_4. We use β_3=-1, β_4=-1, $\theta_0 = -0.5$, $\theta_1 = -0.5$ and $\theta_2 = 0.1$ so MNAR assumption is valid here. Lines as in Figure 8.12.

8.6   Application

8.6.1   Sensitivity of the IPW: Two timepoints example

Figure 8.14: The 95% CI of β_3 and β_4 under MAR: Two timepoints example

In this section we use the two timepoints data as introduced previously in this chapter in Section 8.4.1. Similar to Chapter 7 work, the first plot shows confidence intervals for the group 0 mean (Treatment A group), as the assumed value of θ_2 changes. The horizontal line is the estimate under MAR. The second plot is the same for the Treatment B group (group 1). The third plot is the difference, which is the treatment effect. The upper horizontal line is at zero. This time there seems to be a lot of sensitivity to θ_2, which can lead to significantly positive or significantly negative treatment effect estimates. We can conclude from these results that the IPW method is much more sensitive than maximum likelihood to misspecification.

Sensitivity of the IPW method: Three timepoints example

In this section we use the three timepoints data (schizophrenia) used in Section 8.4.2. In group 1 (Treatment 1) there were 85 patients, but only 41 patients provided a response at time 3 and the other 44 patients dropped out which is equivalent to 51.8% missing rate. There were 88 patients in group 2 receiving Treatment 2, but only 29 patients provided a response at time 3 and the other 59 patients dropped out which is equivalent to 67% missing. Of 345 patients receiving Treatment 3 (group 3), only 199 provided a response at time 3 and the other 146 dropped out which is equivalent to 42% missing. Hence the highest dropout rate is in treatment group 2 and the lowest in group 3. We summarise the dropout percentages for each treatment group in Table 8.10.

The top-left plot in Figure 8.15 provides a 95% CI for the differences in means for group 1 and 3 at time 2 (β_6 - β_4). The down-left is for (β_6 - β_5), the top-right is for (β_9 - β_7) and finally the down-right is for (β_9 - β_8).

Table 8.10: Summary of dropout percentages for treatment groups

	Group 1	Group 2	Group3
Time 2	24%	36%	20%
Time 3	52%	67%	42%

Figure 8.15 looks quite different to the equivalent plot produced for the LME in Chapter 7. There is a much greater effect of θ_2 in Figure 8.15 than in Figure 7.11. In Figure 8.15, θ_2 goes from +/- 0.004, but in Figure 7.11 it is +/- 0.04. The reason is as mentioned before that the IPW method is much more sensitive than LME method. However, similar to the LME plot, Figure 8.15 indicates that the differences in means of treatment groups taken in pairs have wider 95% CIs at time 3 than the CIs at time 2. As mentioned in Chapter 7, the results are slightly different from Chapter 2 (Table 2.19) because we do not estimate at time 1.

Figure 8.15: The 95% CI for the differences in means for groups in the schizophrenia data between time 1 and 2 and between time 2 and 3.
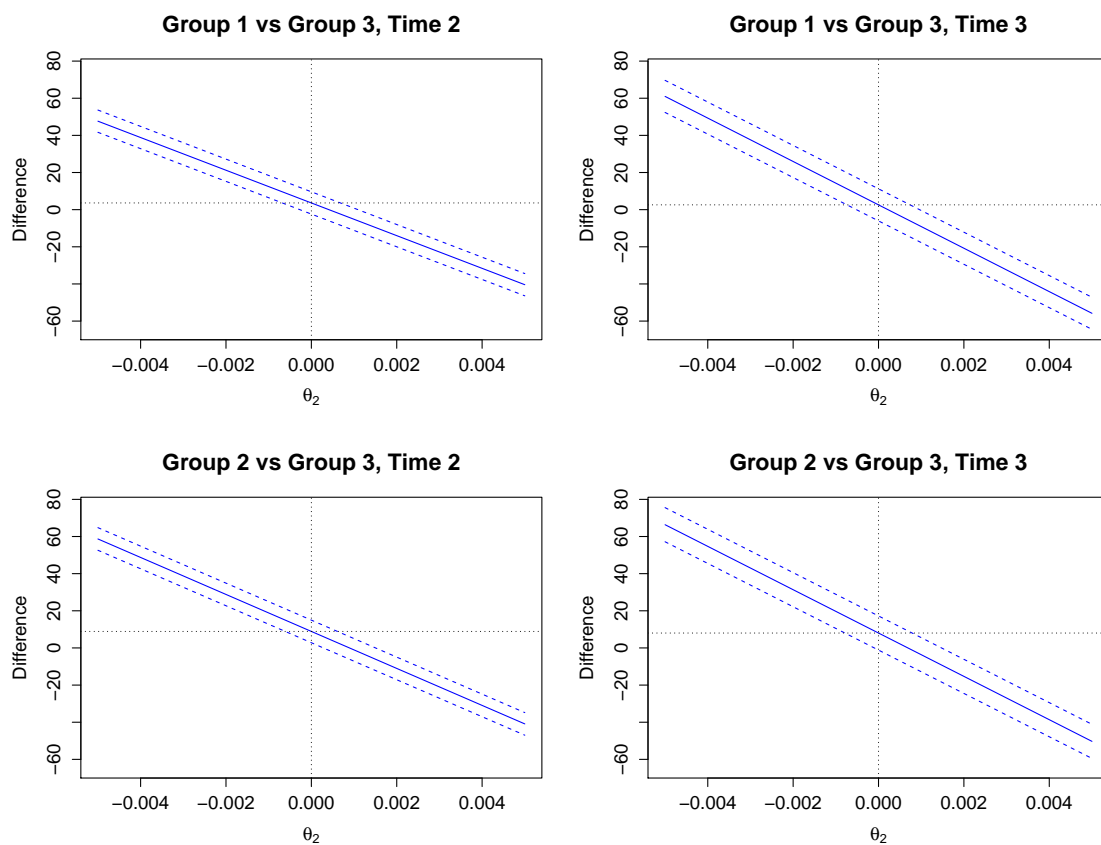
8.7 Conclusion

In this chapter we investigated the effect of local misspecification on dropout probability and IPW estimates. We started with the estimate of the dropout probability (first part) and then we used these estimates to calculate the IPW estimates of regression effects (second part). The main objective is to show that the Copas and Eguchi method to estimate the misspecification is working for the IPW method. We showed how the IPW method is more sensitive to the misspecification by comparing the results obtained here with the previous chapter for the LME method.

Chapter 9

Conclusion and Further Work

9.1 Summary of the Thesis

This last chapter summarizes what is discussed in this thesis about the dropout modeling investigation. Some methods used to handle longitudinal data with dropout were presented. We used the dropout models introduced by Little and Rubin [2002]. The dropout models can be one of the following types: Missing Completely at Random (MCAR), Missing at Random (MAR), Missing Not at Random (MNAR) and Shared Parameter (SP).

The following six general strategies were considered for handling missing data : 1) Complete Case analysis (CC), 2) Observed data analysis (Obs), 3) Inverse Probability Weighted Estimating Equations with Missing at Random Assumption (IPW), 4) Linear Mixed Effect models (LME), 5) Linear Increment models (LI) and 6) Last Observation Carried Forward (LOCF). In Chapter 2, we estimated the parameters of the longitudinal model under MCAR, MAR, MNAR and SP for both simulated data and real data assuming that there are only either two or three scheduled measurements times, for simplicity. We found that all methods work under the MCAR model as expected. Also, the LI method gave consistent estimates under the SP model. The IPW and LME gave consistent estimates under MAR, while no method worked under MNAR.

Then, by deriving the so called least false values, we investigated the consequences of misspecifying the missingness mechanism. We described this method in detail for the Linear Increment method (LI) in Chapter 3. Then, in Chapter 4, we applied the same procedure to investigate the IPW method. In Chapters 5 and 6 we investigated the CC and LME methods respectively. In all cases we looked at SP, MAR and MNAR droput. In order to calculate the least false values, we adopted the procedures used by Ho et al. [2012] to our setting. We gave closed form expressions to calculate the least false values β_3^* and β_4^* for LI, CC and LME methods. For IPW we could provide a closed form for β_3^* under SP, MAR and MNAR while for β_4^* we failed to find closed form under MNAR and we used a numerical calculation instead.

The knowledge of the least false values allowed us to conduct sensitivity analysis which was illustrated in Chapters 7 and 8. Using a likelihood approach, Copas and Eguchi [2005] gave a formula to estimate the bias under the misspecification. The LME is a likelihood based method, and this idea also was adapted for the IPW estimating equation approach. We compared the results found by using Copas and Eguchi method with the results obtained by our method. We concluded that the IPW method is much more sensitive to misspecification than the LME method. Also, we applied the Copas and Eguchi method and our method to estimate the bias for the real data examples.

9.2 Further Work

In this section we consider some limitations of our work and indicate how it might be extended.

9.2.1 Generalization to more than one continuous covariate

Throughout this thesis we assumed the model to have one covariate but we can also generalise this to have more than one covariate. First, it is straightforward to show that the methods that work for one covariate still work for any number. In particular linear increments under shared parameter dropout, and IPW and LME under MAR all give consistent estimates however many covariates there are. To illustrate other situations we consider linear increments under MAR dropout but now with two independent standard normal distributed covariates x_1 and x_2.

As before we will let $D_i = Y_{2i} - Y_{1i}$. We assume the correct model for responses, i.e.:

$$E[Y_{1i}] = \beta_1 + \beta_2 x_{i1} + \beta_3 x_{i2} \text{ and } E[Y_{2i}] = \beta_4 + \beta_5 x_{i1} + \beta_6 x_{i2}.$$

So for fully observed data

$$E[D_i] = (\beta_4 - \beta_1) + (\beta_5 - \beta_2)x_{i1} + (\beta_6 - \beta_3)x_{i2} = \gamma_1 + \gamma_2 x_{i1} + \gamma_3 x_{i2} \quad \text{say.}$$

The parameter estimates are available in closed form.

$$\hat{\gamma_3} = \frac{\left(\sum R_i D^{(M)} x_{2i}^{(M)}\right)\left(\sum R_i (x_{1i}^{(M)})^2\right) - \left(\sum R_i D^{(M)} x_{1i}^{(M)}\right)\left(\sum R_i x_{1i}^{(M)} x_{2i}^{(M)}\right)}{\left(\sum R_i (x_{1i}^{(M)})^2\right)\left(\sum R_i (x_{2i}^{(M)})^2\right) - \left(\sum R_i x_{1i}^{(M)} x_{2i}^{(M)}\right)^2}$$

$$\hat{\gamma_2} = \frac{\left(\sum R_i D^{(M)} x_{1i}^{(M)}\right)\left(\sum R_i (x_{2i}^{(M)})^2\right) - \left(\sum R_i D^{(M)} x_{2i}^{(M)}\right)\left(\sum R_i x_{1i}^{(M)} x_{2i}^{(M)}\right)}{\left(\sum R_i (x_{1i}^{(M)})^2\right)\left(\sum R_i (x_{2i}^{(M)})^2\right) - \left(\sum R_i x_{1i}^{(M)} x_{2i}^{(M)}\right)^2}$$

$$\hat{\gamma}_1 = \frac{\overline{RD}}{\overline{R}} - \hat{\gamma}_2 \frac{\overline{Rx_1}}{\overline{R}} - \hat{\gamma}_3 \frac{\overline{Rx_2}}{\overline{R}}.$$

The superscript (M) is used to indicate that the data are mean corrected, based on the means for the relevant observed data.

As sample size increases these converge to

$$\gamma_3^* = \frac{E[RD^{(M)}x_2^{(M)}]E[R(x_1^{(M)})^2] - E[RD^{(M)}x_1^{(M)}]E[Rx_1^{(M)}x_2^{(M)}]}{E[R(x_1^{(M)})^2]E[R(x_2^{(M)})^2] - \left(E[Rx_1^{(M)}x_2^{(M)}]\right)^2} \tag{9.1}$$

$$\gamma_2^* = \frac{E[RD^{(M)}x_1^{(M)}]E[R(x_2^{(M)})^2] - E[RD^{(M)}x_2^{(M)}]E[Rx_1^{(M)}x_2^{(M)}]}{E[R(x_1^{(M)})^2]ER(x_2^{(M)})^2] - \left(E[Rx_1^{(M)}x_2^{(M)}]\right)^2} \tag{9.2}$$

$$\gamma_1^* = \frac{E[RD]}{E[R]} - \gamma_2^* \frac{E[Rx_1]}{E[R]} - \gamma_3^* \frac{E[Rx_2]}{E[R]}. \tag{9.3}$$

The same ideas of Chapter 3 can be used to obtain these. For example consider $E[R]$. Following the approach presented in Chapter 3, as we know $E[R] = E[E[R|x_1, x_2]]$, thus we have to calculate $E[R|x_1, x_2]$ first, then we can find $E[R]$. Note that under MAR, R does not depend on ϵ_2 as it depends only on x_1, x_2, U and ϵ_1. Also, we assume that x_1 is independent of x_2. We have

$$\begin{aligned}
E[R|x_1, x_2] &= \int P(R = 1|x_1, x_2, \epsilon_1, U)f(\epsilon_1, U)\,\mathrm{d}\epsilon_1\,\mathrm{d}U \\
&= \int \mathrm{expit}\{\theta_0 + \theta_1 Y_1\}f(\epsilon_1, U)\,\mathrm{d}\epsilon_1\,\mathrm{d}U \\
&= \int \mathrm{expit}\{\theta_0 + \theta_1(\beta_1 + \beta_2 x_1 + \beta_3 x_2) + \theta_1(U + \epsilon_1)\}f(\epsilon_1, U)\,\mathrm{d}\epsilon_1\,\mathrm{d}U \\
&= \int \mathrm{expit}\{K_1 + \theta_1 w_1\}f(w_1)\,\mathrm{d}w_1, \text{ say,}
\end{aligned}$$

where

$$\begin{aligned}
w_1 &= \epsilon_1 + U \tag{9.4} \\
K_1 &= \theta_0 + \theta_1(\beta_1 + \beta_2 x_1 + \beta_3 x_2). \tag{9.5}
\end{aligned}$$

We now use an approximation of the expit to the cumulative normal. Therefore:

$$E[R|x_1, x_2] \approx \int \Phi\{c(K_1 + \theta_1 w_1)\}f(w_1)\,\mathrm{d}w_1.$$

Note that w_1 is normally distributed with mean 0 and variance $\sigma_{w_1}^2$, i.e $w_1 \sim N(0, \sigma_{w_1}^2)$, where $\sigma_{w_1}^2 = \sigma_{\epsilon_1}^2 + \sigma_U^2$, which allows us to replace $f(w_1)$ with $\phi(w_1; 0, \sigma_{w_1}^2)$ and

$$E[R|x_1, x_2] \approx \int_{-\infty}^{\infty} \Phi\{c(K_1 + \theta_1 w_1)\}\phi(w_1; 0, \sigma_{w_1}^2)\,\mathrm{d}w_1.$$

Thus,

$$
\begin{aligned}
E[R|x_1, x_2] &\approx \Phi\left[\frac{cK_1}{\sqrt{1 + c^2\theta_1^2\sigma_{w_1}^2}}\right] \\
&= \Phi\left[\frac{c(\theta_0 + \theta_1(\beta_1 + \beta_2 x_1 + \beta_3 x_2))}{\sqrt{1 + c^2\theta_1^2\sigma_{w_1}^2}}\right] \\
&= \Phi(A_1 + A_2 x_1 + A_3 x_2) \quad\quad\quad (9.6)
\end{aligned}
$$

where we have used equation (9.5) to replace K_1 and we have defined $A_1 = \frac{c(\theta_0 + \theta_1\beta_1)}{\sqrt{1 + c^2\theta_1^2\sigma_{w_1}^2}}$, $A_2 = \frac{c(\theta_1\beta_2)}{\sqrt{1 + c^2\theta_1^2\sigma_{w_1}^2}}$ and $A_3 = \frac{c(\theta_1\beta_3)}{\sqrt{1 + c^2\theta_1^2\sigma_{w_1}^2}}$.

Now we can find $E[R]$ by integrating the above expectation in (9.6) over x_1 then x_2. Assume that $x_1 \sim N(0, \sigma_{x_1}^2)$ independent of $x_2 \sim N(0, \sigma_{x_2}^2)$ so

$$
\begin{aligned}
E[R] &= E_{x_1, x_2}[E[R|x_1, x_2]] \\
&\approx \int\int \Phi(A_1 + A_2 x_1 + A_3 x_2)\phi(x_1; 0, \sigma_{x_1}^2)\phi(x_2; 0, \sigma_{x_2}^2)\mathrm{d}x_1\mathrm{d}x_2 \\
&= \int \Phi\left[\frac{A_1 + A_2 x_1}{\sqrt{1 + (A_3\sigma_{x_2})^2}}\right]\phi(x_1; 0, \sigma_{x_1}^2)\mathrm{d}x_1 \\
&= \int \Phi[A_4 + A_5 x_1]\phi(x_1; 0, \sigma_{x_1}^2)\mathrm{d}x_1 \\
&= \Phi\left[\frac{A_4}{\sqrt{1 + (A_3\sigma_{x_2})^2}\sqrt{1 + (A_5\sigma_{x_1})^2}}\right]. \quad\quad (9.7)
\end{aligned}
$$

where $A_4 = A_1/\sqrt{1 + (A_3\sigma_{x_2})^2}$ and $A_5 = A_2/\sqrt{1 + (A_3\sigma_{x_2})^2}$.

The same ideas can be used to obtain the other expressions and hence the least false values. This extends in principle to any number of $N(0, 1)$ covariates, though of course the computations become very tedious.

9.2.2 Generalization to categorical covariate

Most of the theory was based on $N(0, 1)$ covariates but in the real data there was a binary treatment indicator. In Chapters 2, 7 and 8 we used the two timepoints real data example. There are 422 subjects, assigned to either Treatment A or B. Treatment A is associated with treatment effect $x=1$ and treatment B is when $x=0$. Then, at time 2, the mean of the group receiving Treatment B is β_3 and the mean of the group receiving Treatment A is $\beta_3 + \beta_4$.

Thus, to calculate the least false value for categorical data, we estimated θ from the combined data. Because x is binary we can do one group at a time. We set the coefficients

of x to zero, and it will not matter what σ_x is. Simply we assumed β_2 and β_4 are zero. We estimated β_1 and β_3 from the means of the data, one group at a time and with different values (but common θ) for the different groups. The same idea will generalise beyond binary covariates to more general categorical ones: Simply consider one group at a time, as in the previous section. However, the work will become increasingly tedious as the number of categories increase or the number of covariates increases.

9.2.3 Generalization to multiple timepoints

The least false calculations can be expanded to include more than two timepoints. In Chapters 3, 4, 5, and 6 we found the least false values for the two timepoints example. Then in Chapters 7 and 8 we applied the approach used for two timepoints data in the case of three timepoints data by breaking down the three timepoints data to get two data sets at each timepoint. We used treatment 3 as the baseline because it is the largest group. We split this data into four separate data sets, by comparing treatment group 1 with treatment group 3 at time 2 and did the same at time 3. Then we compared treatment group 2 with group 3 at time 2 and again at time 3. Thus one can calculate the least false values for a data set with more than two timepoints using a similar approach of taking the data as subgroups.

This strategy might have some complications, however, if we will apply it for multiple timepoints as more than three timepoints. For studies with long follow up periods, the proportion of individuals with missing data can be substantial and clearly the proportion dropped out increases with time. That is in the case of multiple timepoints the sample size decreases from one timepoint to the next one and at the final timepoint we might have a very large dropout rate or in other words few observations remaining in the data set. This loss of information certainty will impact the estimation efficiency. For example, suppose there are 10 measurements, with 10% of missing data on each measurement. Suppose, further, that missingness on the different measurements is independent. Then, the estimated percentage of incomplete observations is as high as 60% and we would not recommend analysis of data with such a high missing rate.

We tried to use the existing routines in the R code to manage more than three timepoint data. The code can be used as it is to estimate any data set using either the Obs data analysis or the LME methods. For CC, LI and LOCF method the R code can be easily adapted to estimate the parameters of four timepoints and then to five timepoints and so on. Each data set needs separate R code to estimate its parameters using these methods. However, for the IPW method the R code used to estimate the data is complicated and depends on many sub-functions each designed either for two timepoints or three timepoints. These functions can be easily changed to give the estimate for more than three timepoints. Many other works simulate data for more than three timepoints and find the

estimates under the IPW method, see for example Rotnitzky et al. [1998].

9.3 Recommendations

9.3.1 Choice of the method

We also can give a recommendation for the method to be chosen in an analysis by comparing these methods in their estimation efficiency. We can conclude that the LOCF is not recommended because it does not give a consistent estimate under any dropout model. Although the CC and Obs analysis methods are easy to apply, they suffer from loss of information and bias issues which make them again not recommended, except in the case of MCAR, where CC and Obs give consistent estimate but with quite large standard errors. The LI works under SP and MCAR. The IPW and LME work under MAR and MCAR. If the dropout model is MAR, we would recommend LME because the IPW is more sensitive to the dropout and gives large standard errors as we have seen in Chapters 2 and 8. In conclusion, since the MAR dropout model is more common than the others then the LME method can be the most recommended method for the analysis of longitudinal data with dropout.

9.3.2 The variance

Furthermore, in Chapter 3 we investigated the effect of the variance σ_U^2 on the limiting values γ_1^* and γ_2^* using simulated data. We found that changing the variance parameters and keeping the dropout percentage at about 50% has a remarkable effect on the bias. As σ_U (and consequentially σ_{ϵ_1} and σ_{ϵ_2}) increases, the limiting values γ_1^* and γ_2^* go further from the true value ($\gamma_1^G = 1, \gamma_2^G = 1$) and hence the absolute bias increases, which means that the large error variances imply poor results. Similarly, in Chapter 4, we studied the effect of the variance σ_U^2 on the limiting values β_3^* and β_4^*. As for Chapter 3, changing the variance parameters and keeping the same dropout percentage at about 50% has a remarkable effect on the bias. It was clear that as σ_u (and consequentially σ_{ϵ_1} and σ_{ϵ_2}) increases, the limiting values β_3^* and β_4^* go further from the true value ($\beta_3^G = -1, \beta_4^G = -1$) and hence the bias increases, which means that the large error variances imply poor results. In short we can conclude that the more variability, the more bias. This is an area for further work.

9.3.3 Discussing the dropout rate

The dropout rate increases from the current measurement to the upcoming timepoint, because we assume monotone dropout. That is at the baseline we have full data and then at next timepoint some observations are lost and the missingness percentage will increase

among the following timepoints. For example at the baseline we have 100% of the observations then at time two we have only 90% which means the dropout rate is 10% at time one, again at time three say 10% of the current observations are missing thus the total dropout is 19% and so on.

We simulated the data under the final dropout rate at 25% and at 50% and we found that the bias is worse when the dropout rate increases so we choose the dropout rate 40% or 50% to make the bias clearer. However at a lower bias rate such as 15% or 10% some of these methods might not show much bias though they do show bias when the dropout rate increases. Some of the methods are more affected by the dropout rate. For example, the LI method is least affected by the dropout rate increasing while the CC and Obs methods are more affected by increasing of the dropout rate. The dropout rate has a moderate effect on IPW and LME. The bias under MNAR and SP models is higher and clear even at the low dropout rate of 25%. Overall our recommendations is not to attempt to analyse data with more than 50% missingness.

All methods rely on untestable assumptions and at high dropout rates the influence of theses assumptions becomes progressively more pronounced.

9.3.4   Choice of the dropout model

Finally, we can recommend the dropout models MAR and SP models as they are well known. Sensitivity analysis is also recommended as we showed in Chapter 7 and 8.

To sum this chapter up, we discussed some natural extensions to the work presented in this thesis and we gave recommendations for the different choices available for the methods and the dropout model.

Appendix A

A.1 The Extended Skew-Normal Distribution (ESN)

The ESN distribution is described in Johnson et al. [1995]. We will consider the definition and notations of the extended skew-normal distribution that was given in Ho et al. [2012]. A random variable w has an extended skew-normal distribution, $ESN(0, \sigma_w^2, \alpha, \nu)$, if it has density:

$$f(w) = \frac{\phi(w; 0, \sigma_w^2)\Phi(\alpha w + \nu)}{\Phi\left(\nu/\sqrt{1 + \alpha^2 \sigma_w^2}\right)} \tag{A.1}$$

where $\phi(.; 0, \Sigma)$ is the normal density with mean 0 and dispersion Σ and $\Phi(.)$ denotes the distribution function of a standard normal. Note that $f(w)$ is a density function, then the integration of $f(w)$ over w equals 1 which implies:

$$\int_w \phi(w; 0, \sigma_w^2)\Phi(\alpha w + \nu)\, \mathrm{d}w = \Phi\left(\nu/\sqrt{1 + \alpha^2 \sigma_w^2}\right). \tag{A.2}$$

The moment generating function (MGF) is:

$$M(t) = E[e^{t\,w}] = \frac{e^{\frac{1}{2}\sigma_w^2\, t^2}\Phi\left[\frac{\alpha\sigma_w^2\, t + \nu}{\sqrt{1 + \alpha^2 \sigma_w^2}}\right]}{\Phi\left[\nu/\sqrt{1 + \alpha^2 \sigma_w^2}\right]}. \tag{A.3}$$

Hence, the expectation of w is:

$$E[w] = M'(t)_{t=0} = \frac{\alpha\sigma_w^2}{\sqrt{1 + \alpha^2 \sigma_w^2}}\frac{\phi(\bar{\nu})}{\Phi(\bar{\nu})}, \tag{A.4}$$

where $\bar{\nu} = \nu(1 + \alpha^2 \sigma_w^2)^{-\frac{1}{2}}$.

The second moment is:

$$E[w^2] = M''(t)_{t=0} = \frac{\sigma_w^2\,\Phi(\bar{\nu}) - \bar{\nu}(\frac{\alpha\,\sigma_w^2}{\sqrt{1 + \alpha^2 \sigma_w^2}})^2\phi(\bar{\nu})}{\Phi(\bar{\nu})}. \tag{A.5}$$

We are interested in integrals that can be written in terms of ESN. In particular,

$$\int \phi(w; 0, \sigma_w^2) \Phi(\alpha w + \nu)\, dw \;\; = \;\; \Phi\left(\nu/\sqrt{1 + \alpha^2 \sigma_w^2}\right) = \Phi(\bar{\nu}) \tag{A.6}$$

$$\begin{aligned}
\int w\, \phi(w; 0, \sigma_w^2) \Phi(\alpha w + \nu)\, dw \;\; &= \;\; \Phi\left(\nu/\sqrt{1 + \alpha^2 \sigma_w^2}\right) E[w] = \Phi(\bar{\nu}) E[w] \\
&= \;\; \Phi(\bar{\nu}) \frac{\alpha \sigma_w^2}{\sqrt{1 + \alpha^2 \sigma_w^2}} \frac{\phi(\bar{\nu})}{\Phi(\bar{\nu})} \\
&= \;\; \frac{\alpha\, \sigma_w^2\, \phi(\bar{\nu})}{\sqrt{1 + \alpha^2 \sigma_w^2}},
\end{aligned} \tag{A.7}$$

and

$$\begin{aligned}
\int w^2\, \phi(w; 0, \sigma_w^2) \Phi(\alpha w + \nu)\, dw \;\; &= \;\; \Phi\left(\nu/\sqrt{1 + \alpha^2 \sigma_w^2}\right) E[w^2] \\
&= \;\; \Phi(\bar{\nu}) E[w^2] \\
&= \;\; \Phi(\bar{\nu}) \frac{\sigma_w^2\, \Phi(\bar{\nu}) - \bar{\nu}\left(\frac{\alpha \sigma_w^2}{\sqrt{1 + \alpha^2 \sigma_w^2}}\right)^2 \phi(\bar{\nu})}{\Phi(\bar{\nu})} \\
&= \;\; \sigma_w^2\, \Phi(\bar{\nu}) - \bar{\nu}\left(\frac{\alpha \sigma_w^2}{\sqrt{1 + \alpha^2 \sigma_w^2}}\right)^2 \phi(\bar{\nu}).
\end{aligned} \tag{A.8}$$

Also we want to find

$$\int w\, e^{tw}\, \Phi[\nu + \alpha w]\, \phi(w)\, dw. \tag{A.9}$$

We know the MGF of w is:

$$\begin{aligned}
E[e^{tw}] \;\; &= \;\; \int e^{tw}\, \Phi[\nu + \alpha w]\, \phi(w)\, dw \tag{A.10} \\
&= \;\; K\Phi\left[\frac{\alpha \sigma_w^2 t + \nu}{\sqrt{1 + \alpha^2 \sigma_w^2}}\right] e^{\frac{1}{2}\sigma_w^2 t^2}, \text{ say.} \tag{A.11}
\end{aligned}$$

We differentiate it with respect to t, thus equation (A.10) will be

$$\frac{\partial}{\partial t} E[e^{tw}]$$

$$= \frac{\partial}{\partial t} \int e^{tw}\, \Phi[\nu + \alpha w]\, \phi(w)\, dw$$

$$= \int \frac{\partial}{\partial t} e^{tw}\, \Phi[\nu + \alpha w]\, \phi(w)\, dw$$

$$= \int w e^{tw} \, \Phi[\nu + \alpha w] \, \phi(w) \, \mathrm{d}w,$$

and this is exactly what we need in equation (A.9). Hence from equation (A.11)

$$
\begin{aligned}
\int w \, e^{tw} \, \Phi[\nu + \alpha w] \, \phi(w) \, \mathrm{d}w \;&=\; \frac{\partial}{\partial t} K \Phi \left[\frac{\alpha \sigma_w^2 t + \nu}{\sqrt{1 + \alpha^2 \sigma_w^2}} \right] e^{\frac{1}{2} \sigma_w^2 t^2} \\
&=\; K \, \sigma_w^2 \, t \, e^{\frac{1}{2} \sigma_w^2 t^2} \Phi \left[\frac{\alpha \sigma_w^2 t + \nu}{\sqrt{1 + \alpha^2 \sigma_w^2}} \right] \\
&+\; K \, e^{\frac{1}{2} \sigma_w^2 t^2} \frac{\alpha \sigma_w^2}{\sqrt{1 + \alpha^2 \sigma_w^2}} \phi \left(\frac{\alpha \sigma_w^2 t + \nu}{\sqrt{1 + \alpha^2 \sigma_w^2}} \right). \quad (A.12)
\end{aligned}
$$

where $K = 1/\Phi[\bar{\nu}]$.

A.2 Calculating Integral of the Form $\int x\phi(ax + b)\phi(x) \, \mathrm{d}x$

To calculate an integration containing two normal densities we use the following:

$$
\begin{aligned}
\int x\phi(ax + b)\phi(x) \, \mathrm{d}x \;&=\; \int \frac{x}{2\pi} e^{-\frac{1}{2}(ax+b)^2} \, e^{-\frac{1}{2}x^2} \, \mathrm{d}x \\
&=\; \int \frac{x}{2\pi} e^{-\frac{1}{2}(ax+b)^2 + x^2} \, \mathrm{d}x.
\end{aligned}
$$

Let $(ax + b)^2 + x^2 = (a^2 + 1)x^2 + 2abx + b^2 = (a^2 + 1)(x - c)^2 + d$, where $c = -\frac{ab}{a^2+1}$ and $d = \frac{b^2}{a^2+1}$. Then,

$$
\begin{aligned}
\int x\phi(ax + b)\phi(x) \, \mathrm{d}x \;&=\; \int \frac{x}{2\pi} \, e^{-\frac{1}{2}((a^2+1)(x-c)^2 + d)} \, \mathrm{d}x \\
&=\; \frac{e^{-\frac{1}{2}d}}{\sqrt{2\pi}\sqrt{a^2 + 1}} \int \frac{x\sqrt{a^2 + 1}}{\sqrt{2\pi}} \, e^{-\frac{1}{2}(a^2+1)(x-c)^2} \, \mathrm{d}x \\
&=\; \frac{e^{-\frac{1}{2}d}}{\sqrt{2\pi}\sqrt{a^2 + 1}} \int x\phi(x; \mu, \sigma_x^2) \, \mathrm{d}x \\
&=\; \frac{e^{-\frac{1}{2}d}}{\sqrt{2\pi}\sqrt{a^2 + 1}} \mu
\end{aligned}
$$

where $\mu = c, \sigma_x^2 = \frac{1}{a^2+1}$.

A.3 How to Choose θ for Different Dropout Models to Keep the Dropout Rate at about 50%.

The SP model is: $\text{expit}(\theta_0^{sp} + \theta_1^{sp} U)$. First we choose $(\theta_0^{sp}, \theta_1^{sp})$ for shared parameter that make the dropout percentage always at about 50%. Note that these θs do not depend on β.

We know that under SP model,

$$
\begin{aligned}
E[R] &= \int P(R = 1|U) f(U)\, \mathrm{d}U \\
&\approx \Phi\left[\frac{c\theta_0^{sp}}{\sqrt{1 + (c\theta_1^{sp}\sigma_U)^2}}\right].
\end{aligned}
$$

As we want to keep the amount of dropout at 50%, thus we should have $E[R] = 0.5$. This implies $\Phi\left[\frac{c\theta_0^{sp}}{\sqrt{1+(c\theta_1^{sp}\sigma_U)^2}}\right] = 0.5$, but $\Phi[.]$ is the cumulative standard normal, then we have to set $\left[\frac{c\theta_0^{sp}}{\sqrt{1+(c\theta_1^{sp}\sigma_U)^2}}\right]$ to zero. To make a fraction equals 0 either the numerator equals 0 or the denominator is large enough comparing to the numerator. We will choose that the numerator equals 0, hence $\theta_0^{sp} = 0$, and we can choose any value for θ_1.

Then under MAR model:

$\text{expit}\,(\theta_0^M + \theta_1^M Y_1)=\text{expit}\,(\theta_0^M + \theta_1^M(\beta_1 + \beta_2 x + U + \epsilon_1))$. Taking only the θ parts $\theta_0^M + \theta_1^M(\beta_1 + \beta_2 x + U + \epsilon_1)$. At $x = 0$, we can set $\theta_0^{sp} = \theta_0^M + \theta_1^M \beta_1$, and $var\,(\theta_1^M(U + \epsilon_1)) = var\,(\theta_1^{sp} U)$. If $\sigma_U^2 = \sigma_{\epsilon_1}^2$, then $\theta_1^M = \frac{\theta_1^{sp}}{\sqrt{2}}$ and $\theta_0^M = \theta_0^{sp} - \theta_1^M \beta_1 = \theta_0^{sp} - \beta_1 \frac{\theta_1^{sp}}{\sqrt{2}}$.

Similarly for MNAR, the model is:

$\text{expit}\,(\theta_0^N + \theta_1^N Y_2)=\text{expit}\,(\theta_0^N + \theta_1^N(\beta_3 + \beta_4 x + U + \epsilon_2))$. Taking only the θ parts: $\theta_0^N + \theta_1^{MN|}(\beta_3 + \beta_4 x + U + \epsilon_2)$. At $x = 0$, set $\theta_0^{sp} = \theta_0^N + \theta_1^N \beta_3$, and $var\,(\theta_1^N(U + \epsilon_2)) = var(\theta_1^{sp} U)$. If $\sigma_U^2 = \sigma_{\epsilon_2}^2$, then $\theta_1^N = \frac{\theta_1^{sp}}{\sqrt{2}}$ thus $\theta_0^N = \theta_0^{sp} - \theta_1^N \beta_3 = \theta_0^{sp} - \beta_3 \frac{\theta_1^{sp}}{\sqrt{2}}$.

Different combinations of θ under different dropout models are shown in Table A.1.

Table A.1: Different combinations of θ

SP (θ_0, θ_1)	MAR (θ_0, θ_1)	MNAR, $\theta_1 = 0$ (θ_0, θ_2)
$(0, \sqrt{0.5})$	(1,0.5)	(0.5,0.5)
$(0, -\sqrt{0.5})$	(-1,-0.5)	(-0.5,-0.5)
$(0, \sqrt{2})$	(2,1)	(1,1)
$(0, -\sqrt{2})$	(-2,-1)	(-1,-1)

A.4 Calculating β_3^* and β_4^* under SP for the IPW Method

Here we will calculate the following expectation in order to find β_3^* and β_4^* as it is defined in the following:

$$\begin{pmatrix} \beta_3^* \\ \beta_4^* \end{pmatrix} = \begin{pmatrix} E[\frac{R}{\pi}] & E[\frac{Rx}{\pi}] \\ E[\frac{Rx}{\pi}] & E[\frac{Rx^2}{\pi}] \end{pmatrix}^{-1} \begin{pmatrix} E[\frac{RY_2}{\pi}] \\ E[\frac{RY_2x}{\pi}]. \end{pmatrix}. \qquad (A.13)$$

We will use θ_0^* and θ_1^* in calculating the aforementioned expectations in the above equation. Thus we need to calculate $E[R/\pi]$, $E[Rx/\pi]$, $E[Rx^2/\pi]$, $E[RY_2/\pi]$ and $E[RY_2x/\pi]$. We start with,

$$\begin{aligned} E[\frac{R}{\pi}] &= E[R\frac{\{1 + e^{\{\theta_0^* + \theta_1^* Y_1\}}\}}{e^{\{\theta_0^* + \theta_1^* Y_1\}}}] \\ &= E[Re^{-\{\theta_0^* + \theta_1^* Y_1\}}] + E[R]. \end{aligned} \qquad (A.14)$$

We have found $E[R]$ in equation (4.16). Thus now we only have to calculate the first part:

$$\begin{aligned} E[Re^{-\{\theta_0^* + \theta_1^* Y_1\}}] &= E[\text{expit}\{\theta_0 + \theta_1 U\}e^{-\{\theta_0^* + \theta_1^* Y_1\}}] \\ &\approx E[e^{-\theta_0^* - \theta_1^* Y_1}\Phi[c(\theta_0 + \theta_1 U)]] \\ &= E[e^{-\theta_0^* - \theta_1^*(\beta_1 + \beta_2 x + U + \epsilon_1)}\Phi[c(\theta_0 + \theta_1 U)]]. \end{aligned}$$

For simplicity we fix x and take the expectation with respect to U and ϵ_1. Rewrite the above expectation as

$$E[Re^{-\{\theta_0^* + \theta_1^* Y_1\}}|x] \approx a_1 E[e^{a_1 U + a_2 \epsilon_1}\Phi[a_3 + a_4 U]], \qquad (A.15)$$

where $a_1 = e^{-\theta_0^* - \theta_1^*(\beta_1 + \beta_2 x)}$, $a_2 = -\theta_1^*$, $a_3 = c\theta_0$ and $a_4 = c\theta_1$. As we know, ϵ_1 is independent of the others, therefore we can say:

$$E[Re^{-\{\theta_0^* + \theta_1^* Y_1\}}|x] \approx a_1 E[e^{a_2 \epsilon_1}] E[e^{a_1 U}\Phi[a_3 + a_4 U]]. \qquad (A.16)$$

Since $\epsilon_1 \sim N(0, \sigma_{\epsilon_1}^2)$, thus we can calculate $E[e^{a_2 \epsilon_1}]$ using the moment generating function (MGF) of the normal distribution, as

$$E[e^{a_2 \epsilon_1}] = e^{-\frac{1}{2}(a_2 \sigma_{\epsilon_1})^2} = C_1, \text{ say.} \qquad (A.17)$$

The rest of equation (A.16) is

$$E[e^{a_1 U}\Phi[a_3 + a_4 U]] = \int \Phi[a_3 + a_4 U]e^{a_1 U}\phi(U; 0, \sigma_U^2)\,\mathrm{d}U.$$

In this case, the integral is now equivalent to formula (A.3) in Appendix A in Chapter 3; and hence this expectation equals the moment generating function of the ESN, which is

given by

$$
\begin{aligned}
E[e^{a_1 U}] &= M_U(a_2) \\
&= \frac{e^{\frac{1}{2}\sigma_U^2 a_2^2}\Phi\left[\frac{a_4^2\sigma_U^2\,a_2+a_3}{\sqrt{1+a_4^2\sigma_U^2}}\right]}{\Phi\left[a_3/\sqrt{1+a_4^2\sigma_U^2}\right]} \tag{A.18} \\
&= a_5 e^{\frac{1}{2}\sigma_U^2 a_2^2}\Phi\left[\frac{a_4^2\sigma_U^2\,a_2+a_3}{\sqrt{1+a_4^2\sigma_U^2}}\right] = C_2,\text{ say.} \tag{A.19}
\end{aligned}
$$

where $a_5 = 1/\Phi[\bar{\nu}]$ and $\bar{\nu} = a_3(1 + a_4^2\sigma_U^2)^{-\frac{1}{2}}$ We will apply the result of equation (A.19) and equation (A.17) to equation (A.16):

$$
E[Re^{-\{\theta_0^* + \theta_1^* Y_1\}}|x] = a_1 C_1 C_2. \tag{A.20}
$$

Then taking the integral over x

$$
\begin{aligned}
E[Re^{-\{\theta_0^* + \theta_1^* Y_1\}}] &= \int_x a_1 C_1 C_2 \phi(x)\,\mathrm{d}x \\
&= C_1 C_2 \int a_1 \phi(x)\,\mathrm{d}x \\
&= C_1 C_2 C_3 \int e^{a_2\beta_2 x}\phi(x)\,\mathrm{d}x \\
&= C_1 C_2 C_3 e^{\frac{1}{2}(a_2\beta_2\sigma_x)^2} \tag{A.21}
\end{aligned}
$$

where $C_3 = e^{-\theta_0^* - \theta_1^*\beta_1}$ Hence equation (A.14) will be

$$
\begin{aligned}
E[\frac{R}{\pi}] &= C_1 C_2 C_3 e^{\frac{1}{2}(a_2\beta_2\sigma_x)^2} + E[R] \\
&\approx C_1 C_2 C_3 e^{\frac{1}{2}(a_2\beta_2\sigma_x)^2} + \Phi\left[\frac{a_3}{\sqrt{1+(a_4\sigma_x^2)^2}}\right]. \tag{A.22}
\end{aligned}
$$

Note that $E[R]$ was calculated in equation (4.16) The final result can be written as:

$$
E[\frac{R}{\pi}] = J_1. \tag{A.23}
$$

As x is independent of R under SP dropout, and it has mean of zero and variance of σ_x^2, then it is expected to get:

$$
E[\frac{Rx}{\pi}] = 0 = J_2,\text{ say} \tag{A.24}
$$

$$
E[\frac{Rx^2}{\pi}] = \sigma_x^2 E[\frac{R}{\pi}] = \sigma_x^2(J_1) = J_3\text{ say.} \tag{A.25}
$$

In the following we will work on $E[RY_2/\pi]$ and $E[RY_2x/\pi]$.

$$E[\frac{RY_2}{\pi}] \quad = \quad E[RY_2 e^{-\{\theta_0^* + \theta_1^* Y_1\}}] + E[RY_2].$$

The second term is

$$\begin{aligned}
E[RY_2|x] \quad &= \quad E[(\beta_3 + \beta_4 x + U + \epsilon_2)\, R] \\
&= \quad a_6 E[R] + E[UR] + E[R]E[\epsilon_2] \\
&= \quad a_6 E[R] + E[UR].
\end{aligned}$$

We have

$$\begin{aligned}
E[UR] \quad &= \quad E[U\,\mathrm{expit}\{\theta_0 + \theta_1 U\}] \\
&\approx \quad E[U\Phi[a_3 + a_4 U]] \\
&= \quad \frac{a_4\,\sigma_U^2\,\phi(\bar{\nu})}{\sqrt{1 + a_4^2\sigma_U^2}}.
\end{aligned}$$

Also

$$E[RY_2|x] \quad \approx \quad a_6 \Phi\left[\frac{a_3}{\sqrt{1 + (a_4\sigma_x^2)^2}}\right] + \frac{a_4\,\sigma_U^2\,\phi(\bar{\nu})}{\sqrt{1 + a_4^2\sigma_U^2}}$$

where $a_6 = \beta_3 + \beta_4 x$. Taking the integral over x

$$\begin{aligned}
E[RY_2] \quad &\approx \quad \left(\Phi\left[\frac{a_3}{\sqrt{1 + (a_4\sigma_x^2)^2}}\right] + \frac{a_4\,\sigma_U^2\,\phi(\bar{\nu})}{\sqrt{1 + a_4^2\sigma_U^2}}\right) \int a_6\phi(x)\,\mathrm{d}x \\
&= \quad C_4 \int (\beta_3 + \beta_4 x)\phi(x)\,\mathrm{d}x \\
&= \quad C_4\beta_3
\end{aligned}$$

where $C_4 = \Phi\left[\dfrac{a_3}{\sqrt{1+(a_4\sigma_x^2)^2}}\right] + \dfrac{a_4\,\sigma_U^2\,\phi(\bar{\nu})}{\sqrt{1+a_4^2\sigma_U^2}}$. Next

$$
\begin{aligned}
E[RY_2 e^{-\{\theta_0^* + \theta_1^* Y_1\}}|x] &= E[(\beta_3 + \beta_4 x + U + \epsilon_2)\,Re^{-\{\theta_0^* + \theta_1^*(\beta_1 + \beta_2 x + U + \epsilon_1)\}}] \\
&= E[(a_6 + U + \epsilon_2)\,\mathrm{expit}\{\theta_0 + \theta_1 U\}e^{-\{\theta_0^* + \theta_1^*(a_7 + U + \epsilon_1)\}}] \\
&\approx E[(a_6 + U + \epsilon_2)\,\Phi\{c(\theta_0 + \theta_1 U)\}e^{-\{\theta_0^* + \theta_1^*(a_7 + U + \epsilon_1)\}}] \\
&= e^{-\{\theta_0^* + \theta_1^* a_6\}} E[(a_6 + U + \epsilon_2)\,\Phi\{c(\theta_0 + \theta_1 U)\}e^{-\{\theta_1^*(U + \epsilon_1)\}}] \\
&= a_1 E[a_6\,\Phi\{c(\theta_0 + \theta_1 U)\}e^{-\{\theta_1^*(U + \epsilon_1)\}}] \\
&+ a_1 E[U\,\Phi\{c(\theta_0 + \theta_1 U)\}e^{-\{\theta_1^*(U + \epsilon_1)\}}] \\
&+ a_1 E[\epsilon_2\,\Phi\{c(\theta_0 + \theta_1 U)\}e^{-\{\theta_1^*(U + \epsilon_1)\}}] \\
&= G_1 + G_2 + G_3,
\end{aligned}
$$

where $a_7 = \beta_1 + \beta_2 x$. Since ϵ_2 is independent of the others terms, then $G_3=0$. The final results for G_1 and G_2 are:

$$
\begin{aligned}
G_1 &= a_5 a_1 C_1 C_2 \\
G_2 &= a_1 C_1 C_5,
\end{aligned}
$$

where C_1 and C_2 as defined previously in this section, and C_5 is

$$
\begin{aligned}
C_5 &= E[U\,\Phi\{c(\theta_0 + \theta_1 U)\}e^{-\{\theta_1^* U\}}] \\
&= E[U\,\Phi\{a_7 + a_8 U)\}e^{-\{\theta_1^* U\}}] \\
&= \int U\,\Phi\{a_7 + a_8 U)\}e^{-\{\theta_1^* U\}}\phi(U; 0, \sigma_U^2)\,\mathrm{d}U \\
&= \frac{\partial}{\partial a_2}a_6\Phi\left[\frac{a_4\sigma_U^2 a_2 + a_3}{\sqrt{1 + a_4^2\sigma_U^2}}\right]e^{\frac{1}{2}\sigma_U^2 a_2^2} \\
&= a_2 a_6 \sigma_U^2 \Phi\left[\frac{a_4\sigma_U^2 a_2 + a_3}{\sqrt{1 + a_4^2\sigma_U^2}}\right]e^{\frac{1}{2}\sigma_U^2 a_2^2} + a_6 \frac{a_4\sigma_U^2}{\sqrt{1 + a_4^2\sigma_U^2}}\phi\left[\frac{a_4\sigma_U^2 a_2 + a_3}{\sqrt{1 + a_4^2\sigma_U^2}}\right]e^{\frac{1}{2}\sigma_U^2 a_2^2} \\
&= a_6 C_6
\end{aligned}
$$

where $C_6 = a_2\sigma_U^2\Phi\left[\frac{a_4\sigma_U^2 a_2 + a_3}{\sqrt{1+a_4^2\sigma_U^2}}\right]e^{\frac{1}{2}\sigma_U^2 a_2^2} + \frac{a_4\sigma_U^2}{\sqrt{1+a_4^2\sigma_U^2}}\phi\left[\frac{a_4\sigma_U^2 a_2 + a_3}{\sqrt{1+a_4^2\sigma_U^2}}\right]e^{\frac{1}{2}\sigma_U^2 a_2^2}$.

Note that we used formula (A.12) in Appendix A in Chapter 3 to calculate the above integral.

Next

$$
\begin{aligned}
E[RY_2 e^{-\{\theta_0^* + \theta_1^* Y_1\}}|x] &= a_5 a_1 C_1 C_2 + a_1 C_1 C_5 \\
&= a_5 C_1 C_2 e^{-\theta_0^* - \theta_1^*(\beta_1 + \beta_2 x)} + (\beta_3 + \beta_4 x)e^{-\theta_0^* - \theta_1^*(\beta_1 + \beta_2 x)}C_1 C_6.
\end{aligned}
$$

Then finally taking the integral over x

$$
\begin{aligned}
E[RY_2 e^{-\{\theta_0^* + \theta_1^* Y_1\}}] &= \int (G_1 + G_2)\phi(x)\,\mathrm{d}x \\
&= \int G_1\phi(x)\,\mathrm{d}x + \int G_2\phi(x)\,\mathrm{d}x \\
&= \int a_5 a_1 C_1 C_2 \phi(x)\,\mathrm{d}x + \int a_1 C_1 C_5 \phi(x)\,\mathrm{d}x \\
&= a_5 C_1 C_2 \int e^{-\theta_0^* - \theta_1^*(\beta_1 + \beta_2 x)}\phi(x)\,\mathrm{d}x + C_1 C_6 \int (\beta_3 + \beta_4 x) e^{-\theta_0^* - \theta_1^*(\beta_1 + \beta_2 x)}\phi(x)\,\mathrm{d}x \\
&= a_5 C_1 C_2 C_3 e^{\frac{1}{2}(a_2\beta_2\sigma_x)^2} + C_1 C_6 \beta_3 C_3 e^{\frac{1}{2}(a_2\beta_2\sigma_x)^2} + C_1 C_6 \beta_4 C_3 \int x e^{a_2\beta_2 x}\phi(x)\,\mathrm{d}x \\
&= C_7 + C_1 C_6 \beta_4 C_3 \int x e^{a_2\beta_2 x}\phi(x)\,\mathrm{d}x \\
&= C_7 + C_1 C_6 \beta_4 C_3 \sigma_x^2 a_2 \beta_2 e^{a_2\beta_2 x}
\end{aligned}
$$

where $C_7 = a_5 C_1 C_2 C_3 e^{\frac{1}{2}(a_2\beta_2\sigma_x)^2} + C_1 C_6 \beta_3 C_3 e^{\frac{1}{2}(a_2\beta_2\sigma_x)^2}$. We will simply say $E[RY_2/\pi] = J_4$.

Similarly,

$$
\begin{aligned}
E[\frac{Rx\,Y_2}{\pi}] &= E[R\,x\,Y_2 e^{-\{\theta_0^* + \theta_1^* Y_1\}}] + E[R\,x\,Y_2] \\
&= G_4 + G_5, \text{ say.}
\end{aligned}
$$

The final results are

$$
G_4 = C_1 C_3 \sigma_x^2 e^{\frac{1}{2}(a_2\beta_2\sigma_x)^2} \left(a_5 C_2 a_2 \beta_2 + \beta_3 C_6 a_2 \beta_2 + \beta_4 C_6 + \beta_4 C_6 (a_2\beta_2)^2 \sigma_x^2 \right),
$$

and

$$
G_5 = \beta_4 \sigma_x^2 \Phi\left[\frac{a_3}{\sqrt{1 + (a_4\sigma_x^2)^2}} \right].
$$

A.5 Calculating β_3^* and β_4^* under MNAR for the IPW Method

Here we attempt to calculate the following expectation in order to find β_3^* and β_4^*, defined by

$$
\begin{pmatrix} \beta_3^* \\ \beta_4^* \end{pmatrix} = \begin{pmatrix} E[\frac{R}{\pi}] & E[\frac{Rx}{\pi}] \\ E[\frac{Rx}{\pi}] & E[\frac{Rx^2}{\pi}] \end{pmatrix}^{-1} \begin{pmatrix} E[\frac{RY_2}{\pi}] \\ E[\frac{RY_2 x}{\pi}] \end{pmatrix}. \tag{A.26}
$$

We will use θ_0^* and θ_1^* in calculating the aforementioned expectations in the above equation thus we need to calculate the following:

$E[R/\pi]$, $E[Rx/\pi]$, $E[Rx^2/\pi]$, $E[RY_2/\pi]$ and $E[RY_2 x/\pi]$.

We start with,

$$E[\frac{R}{\pi}] = E[R\frac{\{1 + e^{\{\theta_0^* + \theta_1^* Y_1\}}\}}{e^{\{\theta_0^* + \theta_1^* Y_1\}}}]$$
$$= E[Re^{-\{\theta_0^* + \theta_1^* Y_1\}}] + E[R]. \qquad (A.27)$$

We have found $E[R]$ in equation (4.40), thus now we only have to calculate the first part,

$$E[Re^{-\{\theta_0^* + \theta_1^* Y_1\}}] = E[\text{expit}\{\theta_0 + \theta_2 Y_2\}e^{-\{\theta_0^* + \theta_1^* Y_1\}}]$$
$$\approx E[e^{-\theta_0^* - \theta_1^* Y_1}\Phi[c(\theta_0 + \theta_2 Y_2)]]$$
$$= E[e^{-\theta_0^* - \theta_1^*(\beta_1 + \beta_2 x + U + \epsilon_1)}\Phi[c(\theta_0 + \theta_2(\beta_3 + \beta_4 x + U + \epsilon_2))]].$$

For simplicity we fix x and take the expectation with respect to U, ϵ_1 and ϵ_2. Rewrite the above expectation as

$$E[Re^{-\{\theta_0^* + \theta_1^* Y_1\}}|x] \approx b_1 E[e^{b_2 U + b_2 \epsilon_1}\Phi[b_3 + b_4 U + b_4 \epsilon_2]], \qquad (A.28)$$

where $b_1 = e^{-\theta_0^* - \theta_1^*(\beta_1 + \beta_2 x)}$, $b_2 = -\theta_1^*$, $b_3 = c(\theta_0 + \theta_2(\beta_3 + \beta_4 x))$ and $b_4 = c\theta_2$.

As we know, ϵ_1 is independent of the others, therefore we can say:

$$E[Re^{-\{\theta_0^* + \theta_1^* Y_1\}}|x] \approx b_1 E[e^{b_2 \epsilon_1}] E[e^{b_2 U}\Phi[b_3 + b_4 U + b_4 \epsilon_2]], \qquad (A.29)$$

Since $\epsilon_1 \sim N(0, \sigma_{\epsilon_1}^2)$, thus we can calculate $E[e^{b_2 \epsilon_1}]$ using the moment generating function (MGF) of the normal distribution, as

$$E[e^{b_2 \epsilon_1}] = e^{-\frac{1}{2}(b_2\sigma_{\epsilon_1})^2} = L_1, \text{ say.} \qquad (A.30)$$

The rest of equation (A.29) is

$$E[e^{b_2 U}\Phi[b_3 + b_4 U + b_4 \epsilon_2]] = \int\int \Phi[b_3 + b_4 U + b_4 \epsilon_2]e^{b_2 U}\phi(\epsilon_2; 0, \sigma_{\epsilon_2}^2)\phi(U; 0, \sigma_U^2)\,\mathrm{d}\epsilon_2\,\mathrm{d}U$$
$$= \int \left(\int \Phi[b_3 + b_4 U + b_4 \epsilon_2]\phi(\epsilon_2; 0, \sigma_{\epsilon_2}^2)\,\mathrm{d}\epsilon_2\right) e^{b_2 U}\phi(U; 0, \sigma_U^2)\,\mathrm{d}U$$
$$= \int \left(\Phi\left[\frac{b_3 + b_4 U}{\sqrt{1 + (b_4\sigma_{\epsilon_2})^2}}\right]\right) e^{b_2 U}\phi(U; 0, \sigma_U^2)\,\mathrm{d}U$$
$$= \int e^{b_2 U}\Phi[b_5 + b_6 U]\phi(U; 0, \sigma_U^2)\,\mathrm{d}U.$$

where $b_5 = \frac{b_3}{\sqrt{1 + (b_4\sigma_{\epsilon_2})^2}}$ and $b_6 = \frac{b_4}{\sqrt{1 + (b_4\sigma_{\epsilon_2})^2}}$.

In this case, the integral is now equivalent to formula (A.3) in Appendix A in Chapter 3; and hence this expectation equals the moment generating function of the ESN, which is

given by

$$
\begin{aligned}
E[e^{b_2 U}] &= M_U(b_2) \\
&= \frac{e^{\frac{1}{2}\sigma_U^2 b_2^2} \Phi\left[\frac{b_6^2 \sigma_U^2 \, b_2 + b_5}{\sqrt{1+b_6^2 \sigma_U^2}}\right]}{\Phi\left[b_5/\sqrt{1+b_6^2 \sigma_U^2}\right]} = L_2, \text{ say.}
\end{aligned} \tag{A.31}
$$

We will apply the result of equation (A.31) and equation (A.30) to equation (A.29):

$$
E[Re^{-\{\theta_0^* + \theta_1^* Y_1\}}|x] = b_1 L_1 L_2. \tag{A.32}
$$

Taking the integral over x:

$$
\begin{aligned}
E[Re^{-\{\theta_0^* + \theta_1^* Y_1\}}] &= \int b_1 L_1 L_2 \phi(x)\,\mathrm{d}x \\
&= L_1 \int b_1 L_2 \phi(x)\,\mathrm{d}x.
\end{aligned}
$$

Noting that b_1, b_3, b_5 and b_6 involve x, to make progress we need to be able to evaluate integrals of the form

$$
\int g_1(x) \frac{\Phi[g_2(x)]}{\Phi[g_3(x)]} \phi(x)\,\mathrm{d}x.
$$

these are completely intractable and hence we are unable to use our method to approximate $E[R/\pi]$. Similar difficulties occur with the other expectations and so we need numerical methods to evaluate the performance of the IPW method under MNAR dropout.

A.6   Attenuation of Slope under Misspecification

In Figure 3.2 we saw that the slope coefficient γ_2^* decreased towards zero under MAR whether θ_1 was positive or negative. We had similar effects for γ_2^* under MNAR in Figure 3.4 as θ_2 changes and for β_4^* under MNAR in Figure 4.3. Intuitively it seems strange that the bias can be in the same direction for any non zero value of the misspecification parameter θ_1 or θ_2.

The reason can be explained in the following simple diagram, which shows four observations over two x values. Suppose missingness depends on Y through a parameter θ, with $\theta = 0$ implying MCAR. Suppose $\theta > 0$. Then we have more dropout at the very lowest Y values, i.e the blue points in the figure. On the other hand if $\theta < 0$ we have disproportionate dropout amongst the very highest Y values, i.e the red points. In both cases we can see by inspection of the plot that the fitted line without either red or blue points, will be flatter, i.e slope closer to zero.

Figure A.1: Plot of simple regression
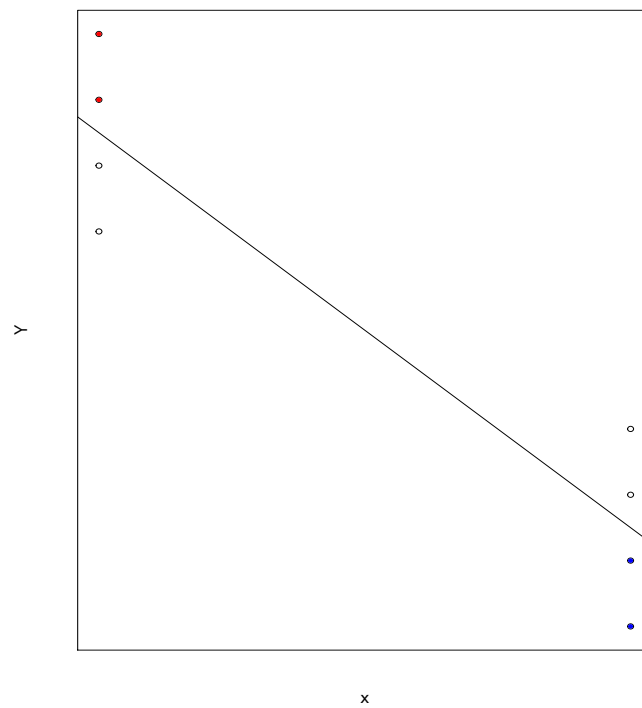
Bibliography

O. Aalen and N. Gunnes. A dynamic approach for reconstructing missing longitudinal data using the linear increments model. *Biostatistics*, 11:453–472, 2010.

A. Afifi and R. Elashoff. Missing observations in multivariate statistics i: Review of the literature. *Journal of the American Statistical Association*, 61:595–604, 1966.

M. L. Bell and D. L. Fairclough. Practical and statistical issues in missing data for longitudinal patient reported outcomes. *Statistical Methods in Medical Research*, 23: 440–59, 2013.

J. Carpenter and J. Bartlett. Simple ad-hoc methods for coping with missing data. Website. URL `http://missingdata.lshtm.ac.uk`. [Accessed: 2016-09-02].

J. R. Carpenter, M. G. Kenward, and S. Vansteelandt. A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society*, 169:571–584, 2006.

R.J. Carrol, D. Rupert, and L.A. Stefanski. *Measurement Error in Nonlinear Models*. Chapman and Hall, London, 1995.

G. Claeskens and NL. Hjort. *Model Selection and Model Averaging*. Cambridge University Press, Cambridge, 2008.

A. Cnaan, N. M. Laird, and P. Slasor. Using the general linear mixed model to analyze unbalanced repeated measures and longitudinal data. *Statistics in Medicine*, 16:2349–2380, 1997.

J. Copas and S. Eguchi. Local model uncertainty and incomplete-data bias (with discussion). *Journal of the Royal Statistical Society*, 67:459–513, 2005.

M. J. Daniels and J. W. Hogan. *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. CRC Press,, New York, 2008.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 39:1–38, 1977.

P. Diggle and M.G. Kenward. Informative drop-out in longitudinal data analysis. *Applied Statistics*, 1:49–93, 1994.

P. Diggle, D. Farewell, and R. Henderson. Analysis of longitudinal data with drop-out: objectives, assumptions and a proposal. *Applied Statistics*, 5:499–550, 2007.

E. Elgmati, D. Farewell, and R. Henderson. A martingale residual diagnostic for longitudinal and recurrent event data. *Lifetime Data Analysis*, 16(1):118–135, 2010.

D. Farewell. *Linear models for censored data. PhD Thesis.* Lancaster University, Lancaster, 2006.

G. M. Fitzmaurice, N. M. Laird, and J. H. Ware. *Applied longitudinal analysis*, volume 998. John Wiley & Sons, 2012.

D. Follmann and M. Wu. An approximate generalized linear model with random effects for informative missing data. *Biometrics*, 51(1):151–168, 1995.

A. Gelman and J. Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press, 2007.

R. Henderson, P. Diggle, and A. Dobson. Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1:465–480, 2000.

M.A. Hernan, E. Lanoy, D. Costagliola, and J.M. Robins. Comparison of dynamic treatment regimes via inverse probability weighting. *Basic & Clinical Pharmacology & Toxicology*, 98:237–242, 2006.

A. Heyting, J. Tolboom, and J. Esser. Statistical handling of dropouts in longitudinal clinical trials. *Statistics in Medicine*, 11:2043–2061, 1992.

W. K. Ho, J. N. S. Matthews, R. Henderson, D. Farewell, and L.R. Rodgers. Dropouts in the ab/ba crossover design. *Statistics in Medicine*, 31:16751687, 2012.

D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47:663–685, 1952.

NL. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions.* Wiley, Chichester, 1995.

I. Kosuke, L. Ying, and S. Aaron. Bayesian and likelihood inference for 2 x 2 ecological tables: An incomplete-data approach. *Political Analysis*, 16:41–69, 2008.

N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38:963–974, 1982.

N. X. Lin, J. Q. Shi, and R. Henderson. Doubly misspecified models. *Biometrika*, 99: 285–298, 2012.

R. J. A. Little and D. B. Rubin. *Statistical analysis with missing data*. Wiley, 2nd edition edition, 2002.

X. De Luna and M. Lundin. Sensitivity analysis of the unconfoundedness assumption with an application to an evaluation of college choice effects on earnings. *Journal of Applied Statistics*, 41(8):1767–1784, 2014.

J. N. Matthews, R. Henderson, D. M. Farewell, W. K. Ho, and L. R. Rodgers. Dropout in crossover and longitudinal studies: Is complete case so bad? *Statistical Methods in Medical Research*, 23:60–73, 2012.

F. Mealli and D. Rubin. Clarifying missing at random and related definitions, and implications when coupled with exchangeability. *Biometrika*, 102:995–1000, 2015.

G. Molenberghs and G. Verbeke. A review on linear mixed models for longitudinal data, possibly subject to dropout. *Statistical Modelling*, 1:235–269, 2001.

G. Molenberghs, H. Thijs, I. Jansen, C. Beunckens, M. G. Kenward, C. Mallinckrodt, and R. J. Carroll. Analyzing incomplete longitudinal clinical trial data. *Biostatistics*, 5(3): 445–464, 2004.

P. M. Philipson, W. K. Ho, and R. Henderson. Comparative review of methods for handling drop-out in longitudinal studies. *Statistics in Medicine*, 27(30):6276–6298, 2008.

J. M. Robins, A. Rotnitzky, and L. P. Zhao. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90:106–121, 1995.

P. R. Rosenbaum. Dilemmas and craftsmanship. In *Design of Observational Studies*, pages 3–20. Springer, 2010.

S. Ross. *A first course in probability*. Prentice Hall press, 2009.

A. Rotnitzky, J. M. Robins, and D. O. Scharfstein. Semiparametric regression for repeated outcomes with nonignorable nonresponse . *Journal of the American Statistical Association*, 93:1321–1339, 1998.

D. B. Rubin. *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York, 1987.

M.D Schluchter. Methods for the analysis of informatively censored longitudinal data. *Statistics of Medicine*, 11:1861–1870, 1992.

S. Seaman, J. Galati, D. Jackson, and J. Carlin. What is meant by missing at random? *Statistical Science*, 28:257–268, 2013.

D. Williams. *Probability with Martingales*. Cambridge University Press, 1991.

M. C. Wu and R. J. Carroll. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 44:175–188, 1988.

H. Zhu, J. G. Ibrahim, and N. Tang. Bayesian sensitivity analysis of statistical models with missing data. *Statistica Sinica*, 24:871–896, 2014.